(FILE 'HOME' ENTERED AT 17:15:57 ON 21 JAN 2007)

FILE 'USPATFULL' ENTERED AT 17:16:09 ON 21 JAN 2007

		E	BARNETT SUSAN W/IN
L1	5	S	E3
		E	MEGEDE J Z/IN
L2	8	S	E4
L3	6	S	L2 NOT L1
		E	ZUR MEGEDE J/IN
L4	2	S	E4
L5	0	S	L4 NOT (L1 OR L2)
		E	LIAN YING/IN
L6	6	S	E3
L7	0	S	L6 NOT (L1 OR L2)
		E	ENGELBRECHT SUSAN/IN
L8	2	S	E3
L9	0	S	L8 NOT (L1 OR L2)
		E	RENSBRUG E J/IN
L10	2	S	E5
L11	0	S	L10 NOT (L1 OR L2)
L12	49677	S	(HIV OR HUMAN IMMUNODEFICIENCY VIRUS)
L13	725	S	L12 AND (CODON-OPTIMIZ?)
L14	697	S	L13 AND (ENV? OR GP160 OR GP140 OR GP120)
L15	154	S	L14 AND (V2)
L16	10	S	L15 AND (DELET? V2 OR V2 DELET?)
L17	1	S	L16 NOT (L1 OR L2)
		E	HAIGWOOD NANCY/IN
L18	14	S	E4
L19	14	S	L18 NOT (L1 OR L2)

FILE 'MEDLINE' ENTERED AT 17:25:24 ON 21 JAN 2007

		E	BARNETT S W/AU
L20	29	S	E3
		E	BARNETT SUSAN W/AU
L21	26	S	E3
L22	55	S	L20 OR L21
L23	49	S	L22 AND (HIV OR HUMAN IMMUNODEFICIENCY VIRUS)
		E	MEGEDE J Z/AU
L24	5	S	E3 OR E4
L25	4	S	L24 NOT L22
		E	ZUR MEGEDE J/AU
L26	23	S	E3 OR E4
L27	11	S	L26 NOT L22
L28	172238	S	(HIV OR HUMAN IMMUNODEFICIENCY VIRUS)
L29	56	S	L28 AND (CODON-OPTIMIZ?)
L30	26	S	L29 AND (ENV? OR GP120 OR GP160 OR GP140)
L31	26	S	L30 NOT (L22)
L32	15712	S	L28 AND (ENV? OR GP160 OR GP140 OR GP120)
L33	312	S	L32 AND (V2)
L34	60	S	L33 AND (DELET?)
L35	20	S	L34 AND (V2 (8W) DELET?)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

Bib Data Sheet

CONFIRMATION NO. 1709

SERIAL NUMBER	FILING OR 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
09/899,575	07/05/2001 RULE	536	1648	PP01631.102 (CHIR-1631/03

APPLICANTS

Jan Zur Megede, San Francisco, CA;
Susan W. Barnett, San Francisco, CA;
Ying Lian, Vallejo, CA;
Susan Engelbrecht, Tygerberg, SOUTH AFRICA;
Estrelita Janse van Rensburg, Tygerberg, SOUTH AFRICA;

**** CONTINUING DATA ***************************

This application is a CIP of 09/610,313 07/05/2000

**** FOREIGN APPLICATIONS ********************

IF REQUIRED, FOREIGN FILING LICENSE GRANTED
** 08/06/2001

Foreign Priority claimed	STATE OR COUNTRY	SHEETS DRAWING	TOTAL CLAIMS	INDEPENDENT CLAIMS
☐ yes ☐ no	CA	114	97	30
35 USC 119 (a-d) conditions met ☐ yes ☐ no ☐ Met after Allowance				
Verified and Acknowledged Examiner's Signature Initials				

ADDRESS

Anne S. Dollard
CHIRON CORPORATION
Intellectual Property - R440
P.O. Box 8097
Emeryville, CA94662-8097

TITLE

Polynucleotides encoding antigenic HIV type C polypeptides, polypeptides and uses thereof

FILING FEE RECEIVED		
4524	FEES: Authority has been given in Paper No. _____ to charge/credit DEPOSIT ACCOUNT No. _____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

AMENDMENTS TO THE CLAIMS

This listing of the claims replaces all prior listings and versions:

1. (canceled).

2. (withdrawn and currently amended): An expression cassette comprising a polynucleotide comprising a sequence encoding an immunogenic Env polypeptide and having at least 90% percent identity to the full-length of the sequence shown in SEQ ID NO:46.

3. (withdrawn and currently amended): The expression cassette of claim 2, comprising a polynucleotide comprising a sequence encoding an immunogenic Env polypeptide and having at least 90% percent identity to the full-length of the sequence shown in SEQ ID NO:47.

4. (withdrawn and currently amended): The expression cassette of claim 3, comprising a polynucleotide comprising a sequence encoding an immunogenic Env polypeptide and having at least 90% percent identity to the full-length of the sequence shown in SEQ ID NO:49 or SEQ ID NO:97.

5 to 37. (canceled).

38. (currently amended): An expression cassette comprising a polynucleotide comprising a sequence encoding an immunogenic Env polypeptide and having at least 90% percent identity to the full-length sequence of the sequence shown in SEQ ID NO:119; SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; SEQ ID NO:124; SEQ ID NO:125; SEQ ID NO:126; SEQ ID NO:127; SEQ ID NO:131; SEQ ID NO:132 or SEQ ID NO:133.

39 to 77. (canceled).

78. (withdrawn): A method of immunization of a subject, comprising,

introducing an expression cassette of claim 38 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

79. (withdrawn): The method of claim 78, wherein said expression cassette is introduced using a gene delivery vector.

80. (withdrawn): The method of claim 79, wherein the gene delivery vector is a non-viral vector.

81. (withdrawn): The method of claim 79, wherein said gene delivery vector is a viral vector.

82. (withdrawn): The method of claim 79, wherein said gene delivery vector is selected from the group consisting of an adenoviral vector, a vaccinia viral vector, an AAV vector, a retroviral vector, a lentiviral vector and an alphaviral vector.

83. (withdrawn): The method of claim 82, wherein said gene delivery vector is a Sindbis-virus derived vector.

84. (withdrawn): The method of claim 82, wherein said gene delivery vector is a cDNA vector.

85. (withdrawn): The method of claim 82, wherein said gene delivery vector is a eukaryotic layered viral initiation system (ELVIS).

86. (withdrawn): The method of claim 79, wherein said composition delivered using a particulate carrier.

87. (withdrawn): The method of claim 79, wherein said composition is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

88. (withdrawn): The method of claim 79, wherein said composition is encapsulated in a liposome preparation.

89. (withdrawn): The method of claim 79, wherein said subject is a mammal.

90. (withdrawn): The method of claim 89, wherein said mammal is a human.

91. (withdrawn): A method of generating an immune response in a subject, comprising:

providing an expression cassette of claim 38,

expressing said polypeptide in a suitable host cell,

isolating said polypeptide, and

administering said polypeptide to the subject in an amount sufficient to elicit an immune response.

92. (withdrawn): A method of generating an immune response in a subject, comprising

introducing into cells of said subject an expression cassette of claim 38, under conditions that permit the expression of said polynucleotide and production of said polypeptide, thereby eliciting an immunological response to said polypeptide.

93. (withdrawn): The method of claim 92, where the method further comprises co-administration of an HIV polypeptide.

94. (withdrawn): The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out before introducing said expression cassette.

95. (withdrawn): The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out concurrently with introducing said expression cassette.

96. (withdrawn): The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out after introducing said expression cassette.

97. (canceled).

SCORE - View Sequence Detail(s) for Applic
09899575

Enter SEQ ID No:
Submit
Enter Application ID No:
Submit
First Sequence
Next Sequence
Previous Sequence
Last Sequence
Convert To Search Format
Go back to Table of Contents Page
Download All Sequences

Here is the list of the requested sequences:

<210> SEQ ID NO 120
<211> LENGTH: 1986
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: gp140mod.TV1.d
<400> SEQUENCE: 120

```
gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc 60
ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac 120
ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc 180
tacgagaccg aggtgcacaa cgtgtgggcc agcccacgcct gcgtgcccac cgaccccaac 240
ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg 300
gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg 360
aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc 420
accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag 480
atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc 540
atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc 600
gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg ccctgctac 660
aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg 720
ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac 780
accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgcccccaac 840
aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac 900
gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc 960
ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag 1020
ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc 1080
ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac 1140
aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg 1200
cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc 1260
tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac 1320
aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg 1380
tacaagtaca aggtggtgga gatcaagccc ctgggcatcg cccccaccaa ggccaagcgc 1440
cgcgtggtgc agcgcgagaa gcgcgccgtg ggcatcggcg ccgtgttcct gggcttcctg 1500
ggcgccgccg gcagcaccat gggcgccgcc agcatcaccc tgaccgtgca ggcccgccag 1560
ctgctgagcg gcatcgtgca gcagcagagc aacctgctga aggccatcga ggcccagcag 1620
cacatgctgc agctgaccgt gtggggcatc aagcagctgc aggcccgcgt gctggccatc 1680
gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg ccgcctgatc 1740
tgcaccaccg ccgtgccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg 1800
gacaacatga cctggatgca gtgggaccgc gagatcagca actacaccgg cctgatctac 1860
aacctgctgg aggacagcca gaaccagcag gagaagaacg agaaggacct gctggagctg 1920
gacaagtgga acaacctgtg gaactggttc gacatcagca actggccctg gtacatctaa 1980
ctcgag                                                            1986
```

performed in any order. In any of the embodiments described herein, the nucleic acid molecules can encode all, some or none of the polypeptides. Thus, one or more or the nucleic acid molecules (*e.g.*, expression cassettes) described herein and/or one or more of the polypeptides described herein can be co-administered in any order and via any administration routes. Therefore, any combination of polynucleotides and/or polypeptides described herein can be used to generate elicit an immune reaction.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1

Generation of Synthetic Expression Cassettes

A. Modification of HIV-1 *Env, Gag, Pol* Nucleic Acid Coding Sequences

The Pol coding sequences were selected from Type C strain AF110975. The Gag coding sequences were selected from the Type C strains AF110965 and AF110967. The Env coding sequences were selected from Type C strains AF110968 and AF110975. These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., *J*

Virol. 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements are inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins. Figures 5 and 6 (SEQ ID Nos: 3, 4, 20 and 21) show the location of some remaining INS in synthetic sequences derived from strains AF110965 and AF110967. The changes made to these sequences are boxed in the Figures. In Figures 5 and 6, the top line depicts a modified sequence of Gag polypeptides from the indicated strains. The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to further remove INS. Thus, when the changes indicated in the boxed regions are made, the resulting sequences correspond to the sequences depicted in Figures 1 and 2, respectively.

The synthetic coding sequences are assembled by methods known in the art, for example by companies such as the Midland Certified Reagent Company (Midland, Texas).

In one embodiment of the invention, sequences encoding Pol-polypeptides are included with the synthetic Gag or Env sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, modified Gag/Env expression cassette. Because synthetic HIV-1 Pol expresses the functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it may be helpful in some instances to inactivate RT and INT functions. Several deletions or mutations in the RT and INT coding regions can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* **9(1):**58-68; Wakefield, J. K.,et al., (1992) *Journal Of Virology* **66(11):**6806-6812; Esnouf, R.,et al., (1995) *Nature Structural Biology* **2(4):**303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* **282(2):**359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* **73** (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **90(13)**:6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* **269(46)**:29279-29287; Goldgur, Y., et

Table D

Name	Seq Id	Description
gp120mod.TV1.delV2	119	synthetic sequence of Env gp120, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.delV2	120	synthetic sequence of Env gp140, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.mut7.delV2	121	synthetic sequence of Env gp140, including V2 deletion and mutation in cleavage site and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV1V2	122	synthetic sequence of Env gp160, including V1/V2 deletion and modified leader derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV2	123	synthetic sequence of Env gp160, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.mut7.delV2	124	synthetic sequence of Env gp160, including V2 deletion; a mutation in cleavage site; and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.tpa1	125	synthetic sequence of Env gp160, TPA1 leader
gp160mod.TV1	126	synthetic sequence of Env gp160, including modified leader sequences derived from wild-type (8_2_TV1_C.ZA) sequences
gp160mod.TV1.wtLnative	127	synthetic sequence of Env gp160, including wild type 8_2_TV1_C.ZA (unmodified) leader
gp140.mod.TV1.tpa1	131	synthetic sequence of Env gp140, TPA1 leader
gp140mod.TV1	132	synthetic sequence of Env gp140, including modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.wtLnative	133	synthetic sequence of Env gp120, including wild type 8_2_TV1_C.ZA (unmodified) leader sequence.

As noted above, Env-encoding constructs can be prepared using any of the full-length of gp160 constructs. For example, a gp140 form (SEQ ID NO:132) was made by truncating gp160 (SEQ ID NO:126) at nucleotide 2064; gp120 was made by truncating gp160 (SEQ ID NO:126) at nucleotide 1551 (SEQ ID NO:126). Additional gp140 and gp120 forms can be made using the methods described herein. One or more stop codons are typically added (*e.g.*, nucleotides 2608 to 2610 of SEQ ID NO:126). Further, the wild-type leader sequence can be modified and/or replaced with other leader sequences (*e.g*, TPA1 leader sequences).

Thus, the polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure. A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (*i.e.* trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be mutated in a variety of ways. In the exemplary constructs described herein (*e.g.*, SEQ ID NO:121 and SEQ ID NO:124), the mutation in the gp120/gp41 cleavage site changes the wild-type amino acid sequence KRRVVQREKR (SEQ ID NO:129) to ISSVVQSEKS (SEQ ID NO:130).

In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated. Exemplary constructs having variable region deletions (V1 and/or V2), V2 deletes were constructed by deleting nucleotides from approximately 499 to approximately 593 (relative to SEQ ID NO:128) and V1/V2 deletes were constructed by deleting nucleotides from approximately 375 to

↓: is the regions for β-sheet deletions

*: is the N-linked glycosylation sites for subtype C TV1 and TV2. Possible mutation (N→ Q) or deletions can be performed.

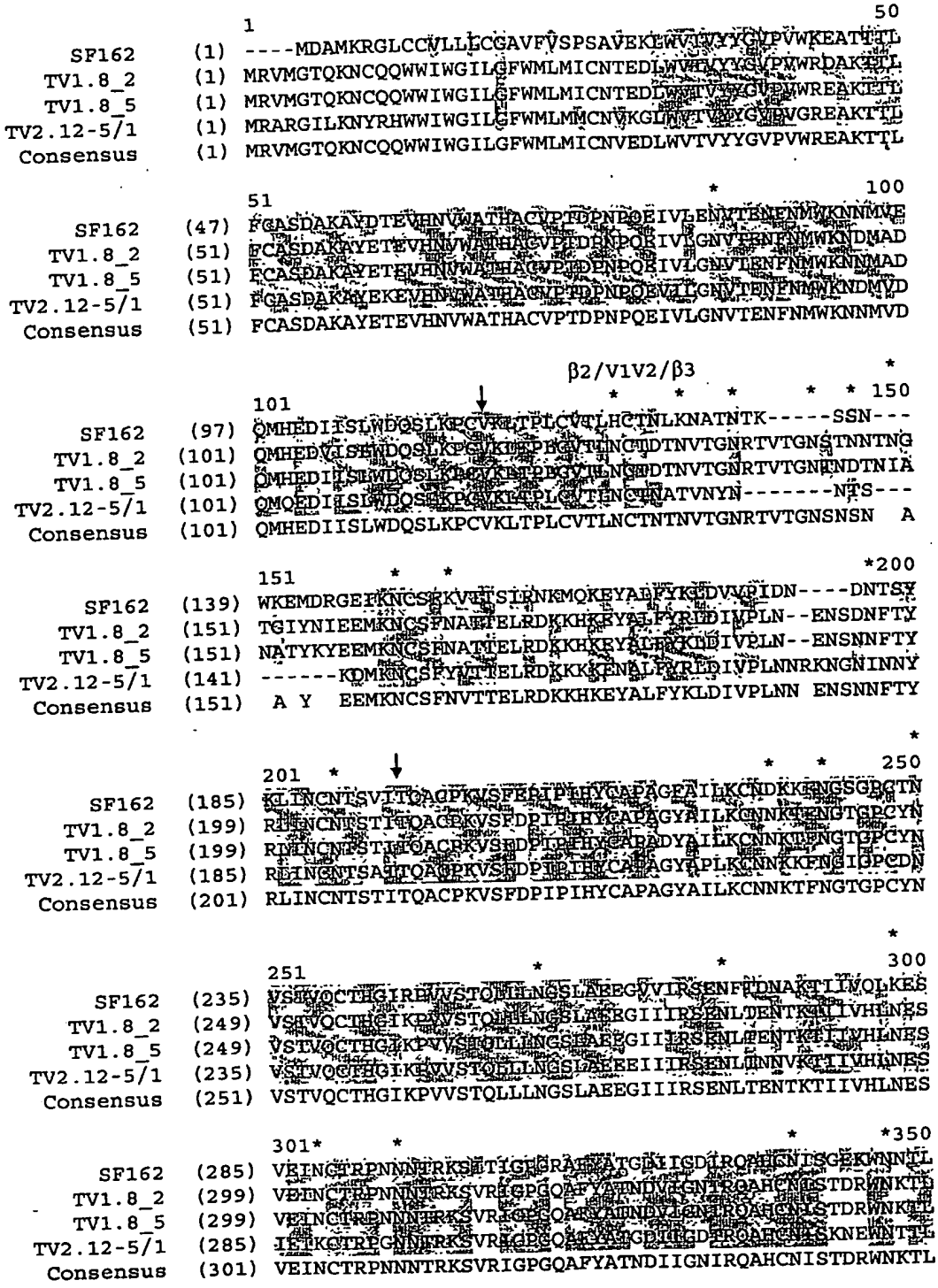


FIGURE 105A

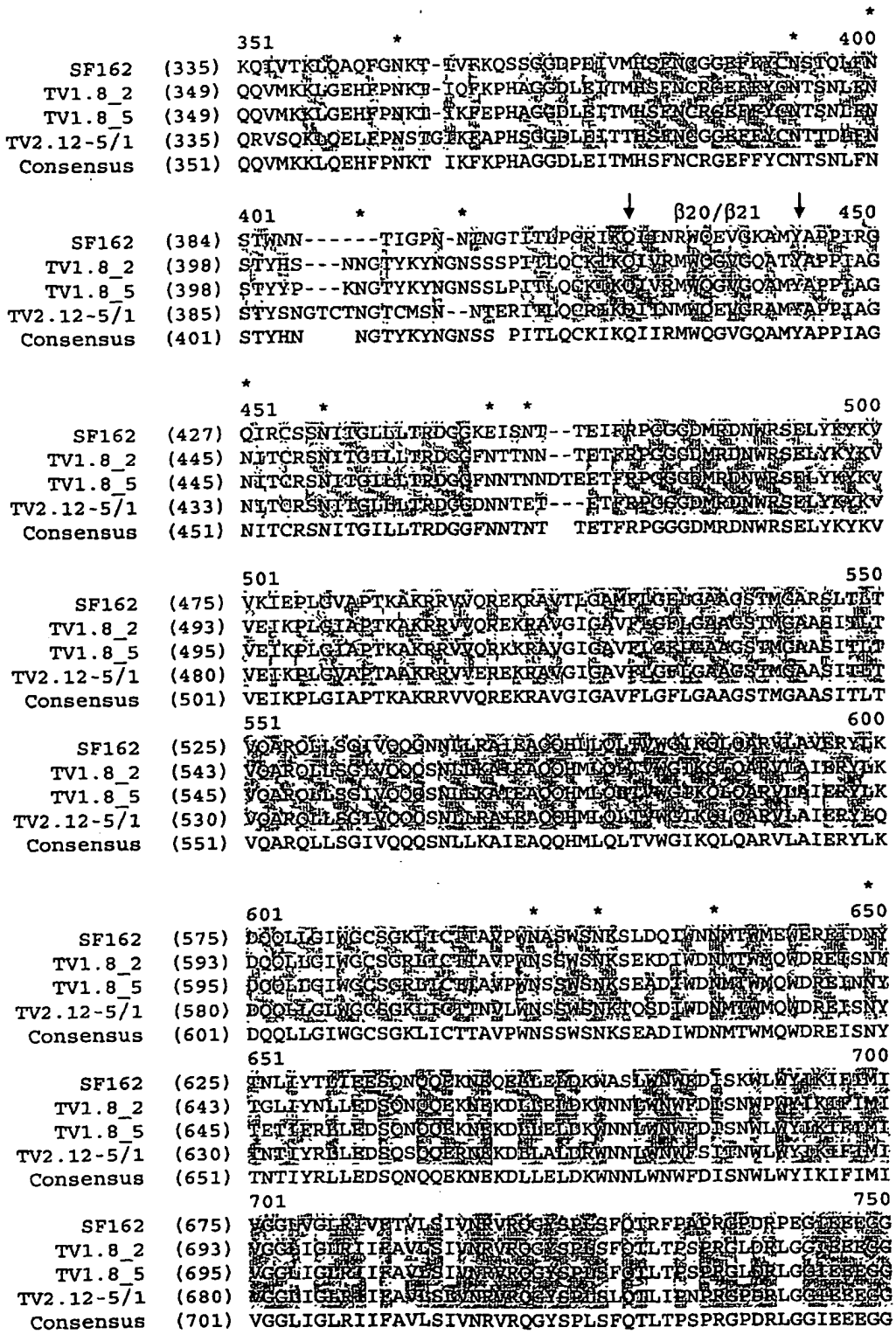


FIGURE 105B

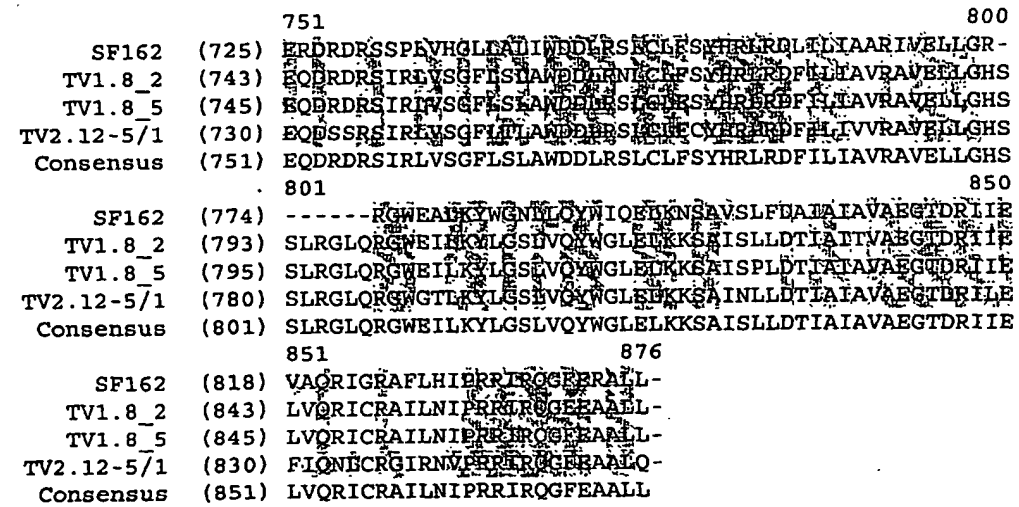

751
800
SF162 (725)
TV1.8_2 (743)
TV1.8_5 (745)
TV2.12-5/1 (730)
Consensus (751) EQDRDRSIRLVSGFLSLAWDDLRSLCLFSYHRLRDFILIAVRAVELLGHS
801
850
SF162 (774)
TV1.8_2 (793)
TV1.8_5 (795)
TV2.12-5/1 (780)
Consensus (801) SLRGLQRGWEILKYLGSLVQYWGLELKKSAISLLDTIAIAVAEGTDRIIE
851
876
SF162 (818)
TV1.8_2 (843)
TV1.8_5 (845)
TV2.12-5/1 (830)
Consensus (851) LVQRICRAILNIPRRIRQGFEAALL

FIGURE 105C

EBI Home Page

EBI

Get Nucleoti

SEQUENCE ANALYSIS

- Help
- General Help
- Formats
- Gaps
- Matrix
- References
- ClustalW Help
- ClustalW FAQ
- Jalview Help
- Scores Table
- Alignment
- Guide Tree
- Colours

ClustalW Results

Results of search	
Number of sequences	14
Alignment score	1075087
Sequence format	Pearson
Sequence type	nt
ClustalW version	1.82
JalView	
Output file	clustalw-20050726-22511744.output
Alignment file	clustalw-20050726-22511744.aln
Guide tree file	clustalw-20050726-22511744.dnd
Your input file	clustalw-20050726-22511744.input

SUBMIT ANOTHER JOB

To save a result file right-click the file link in the above table and choose "Sa
If you cannot see the JalView button, reload the page and check your browse

Scores Table

Sort by Sequence Number View Output File

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	seqid46	97	2	seqid47	144	100
1	seqid46	97	3	seqid119	1473	97
1	seqid46	97	4	seqid120	1986	97
1	seqid46	97	5	seqid121	1986	97
1	seqid46	97	6	seqid122	2397	97
1	seqid46	97	7	seqid123	2529	97
1	seqid46	97	8	seqid124	2529	97

1	seqid46	97	9	seqid125	2613	97
1	seqid46	97	10	seqid126	2616	97
1	seqid46	97	11	seqid127	2616	97
1	seqid46	97	12	seqid131	2052	97
1	seqid46	97	13	seqid132	2073	97
1	seqid46	97	14	seqid133	2073	97
2	seqid47	144	3	seqid119	1473	96
2	seqid47	144	4	seqid120	1986	96
2	seqid47	144	5	seqid121	1986	96
2	seqid47	144	6	seqid122	2397	96
2	seqid47	144	7	seqid123	2529	96
2	seqid47	144	8	seqid124	2529	96
2	seqid47	144	9	seqid125	2613	96
2	seqid47	144	10	seqid126	2616	96
2	seqid47	144	11	seqid127	2616	96
2	seqid47	144	12	seqid131	2052	96
2	seqid47	144	13	seqid132	2073	96
2	seqid47	144	14	seqid133	2073	96
3	seqid119	1473	4	seqid120	1986	99
3	seqid119	1473	5	seqid121	1986	98
3	seqid119	1473	6	seqid122	2397	90
3	seqid119	1473	7	seqid123	2529	99
3	seqid119	1473	8	seqid124	2529	98
3	seqid119	1473	9	seqid125	2613	93
3	seqid119	1473	10	seqid126	2616	99
3	seqid119	1473	11	seqid127	2616	98
3	seqid119	1473	12	seqid131	2052	93
3	seqid119	1473	13	seqid132	2073	99
3	seqid119	1473	14	seqid133	2073	98
4	seqid120	1986	5	seqid121	1986	99
4	seqid120	1986	6	seqid122	2397	93
4	seqid120	1986	7	seqid123	2529	99
4	seqid120	1986	8	seqid124	2529	99
4	seqid120	1986	9	seqid125	2613	95
4	seqid120	1986	10	seqid126	2616	99
4	seqid120	1986	11	seqid127	2616	98
4	seqid120	1986	12	seqid131	2052	95
4	seqid120	1986	13	seqid132	2073	99
4	seqid120	1986	14	seqid133	2073	99
5	seqid121	1986	6	seqid122	2397	92
5	seqid121	1986	7	seqid123	2529	99
5	seqid121	1986	8	seqid124	2529	99
5	seqid121	1986	9	seqid125	2613	94
5	seqid121	1986	10	seqid126	2616	99
5	seqid121	1986	11	seqid127	2616	98
5	seqid121	1986	12	seqid131	2052	94
5	seqid121	1986	13	seqid132	2073	99
5	seqid121	1986	14	seqid133	2073	98
6	seqid122	2397	7	seqid123	2529	100
6	seqid122	2397	8	seqid124	2529	99
6	seqid122	2397	9	seqid125	2613	96
6	seqid122	2397	10	seqid126	2616	99
6	seqid122	2397	11	seqid127	2616	99
6	seqid122	2397	12	seqid131	2052	85
6	seqid122	2397	13	seqid132	2073	89
6	seqid122	2397	14	seqid133	2073	88
7	seqid123	2529	8	seqid124	2529	99
7	seqid123	2529	9	seqid125	2613	96
7	seqid123	2529	10	seqid126	2616	99
7	seqid123	2529	11	seqid127	2616	99

7	seqid123	2529	12	seqid131	2052	91
7	seqid123	2529	13	seqid132	2073	95
7	seqid123	2529	14	seqid133	2073	94
8	seqid124	2529	9	seqid125	2613	96
8	seqid124	2529	10	seqid126	2616	99
8	seqid124	2529	11	seqid127	2616	99
8	seqid124	2529	12	seqid131	2052	91
8	seqid124	2529	13	seqid132	2073	95
8	seqid124	2529	14	seqid133	2073	94
9	seqid125	2613	10	seqid126	2616	96
9	seqid125	2613	11	seqid127	2616	96
9	seqid125	2613	12	seqid131	2052	99
9	seqid125	2613	13	seqid132	2073	95
9	seqid125	2613	14	seqid133	2073	95
10	seqid126	2616	11	seqid127	2616	99
10	seqid126	2616	12	seqid131	2052	96
10	seqid126	2616	13	seqid132	2073	99
10	seqid126	2616	14	seqid133	2073	99
11	seqid127	2616	12	seqid131	2052	96
11	seqid127	2616	13	seqid132	2073	99
11	seqid127	2616	14	seqid133	2073	99
12	seqid131	2052	13	seqid132	2073	96
12	seqid131	2052	14	seqid133	2073	96
13	seqid132	2073	14	seqid133	2073	99

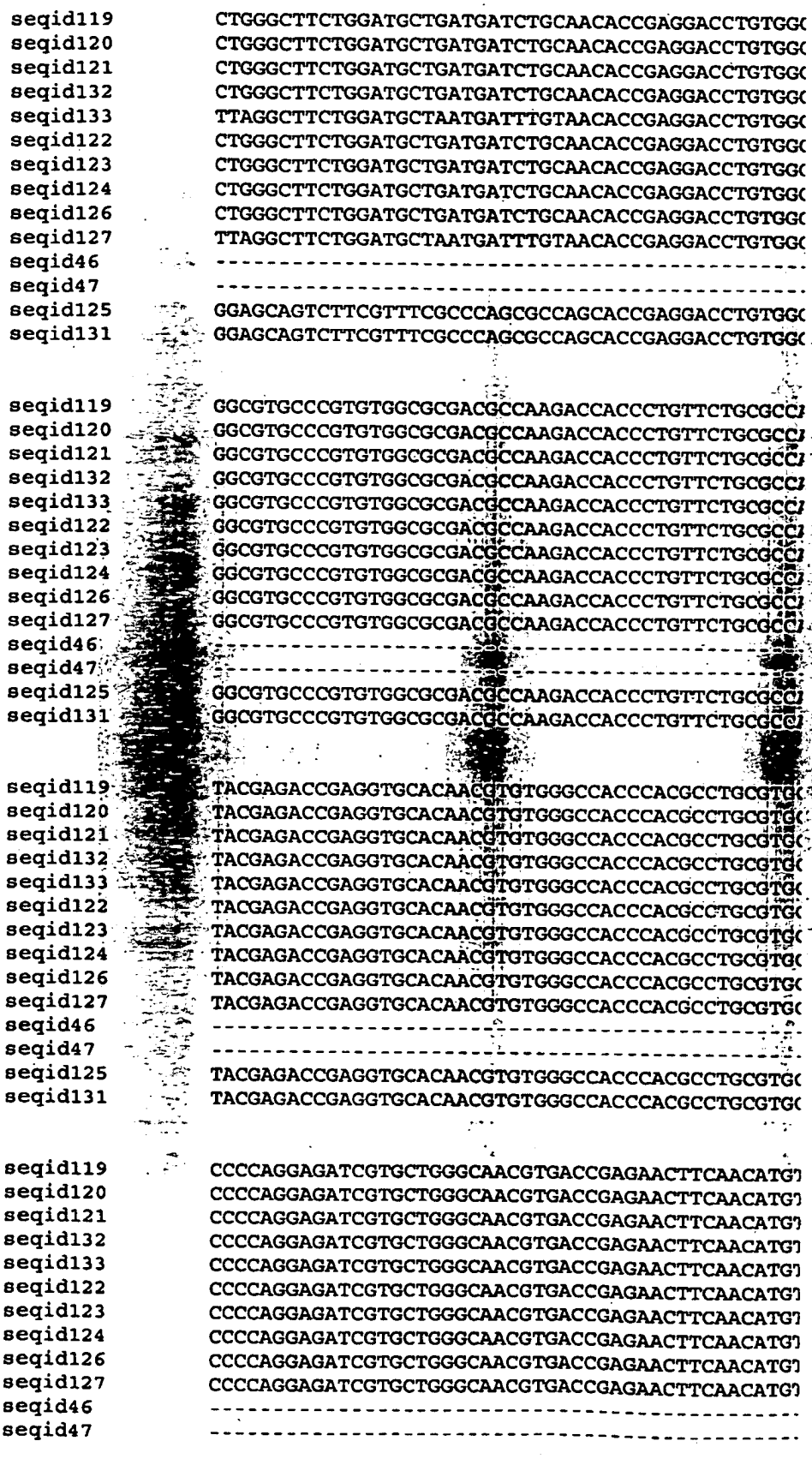
PLEASE NOTE: Some scores may be missing from the above table if the alig output

Sort by | Sequence Number | View Output File | S

Alignment

Show Colors | View Alignment File

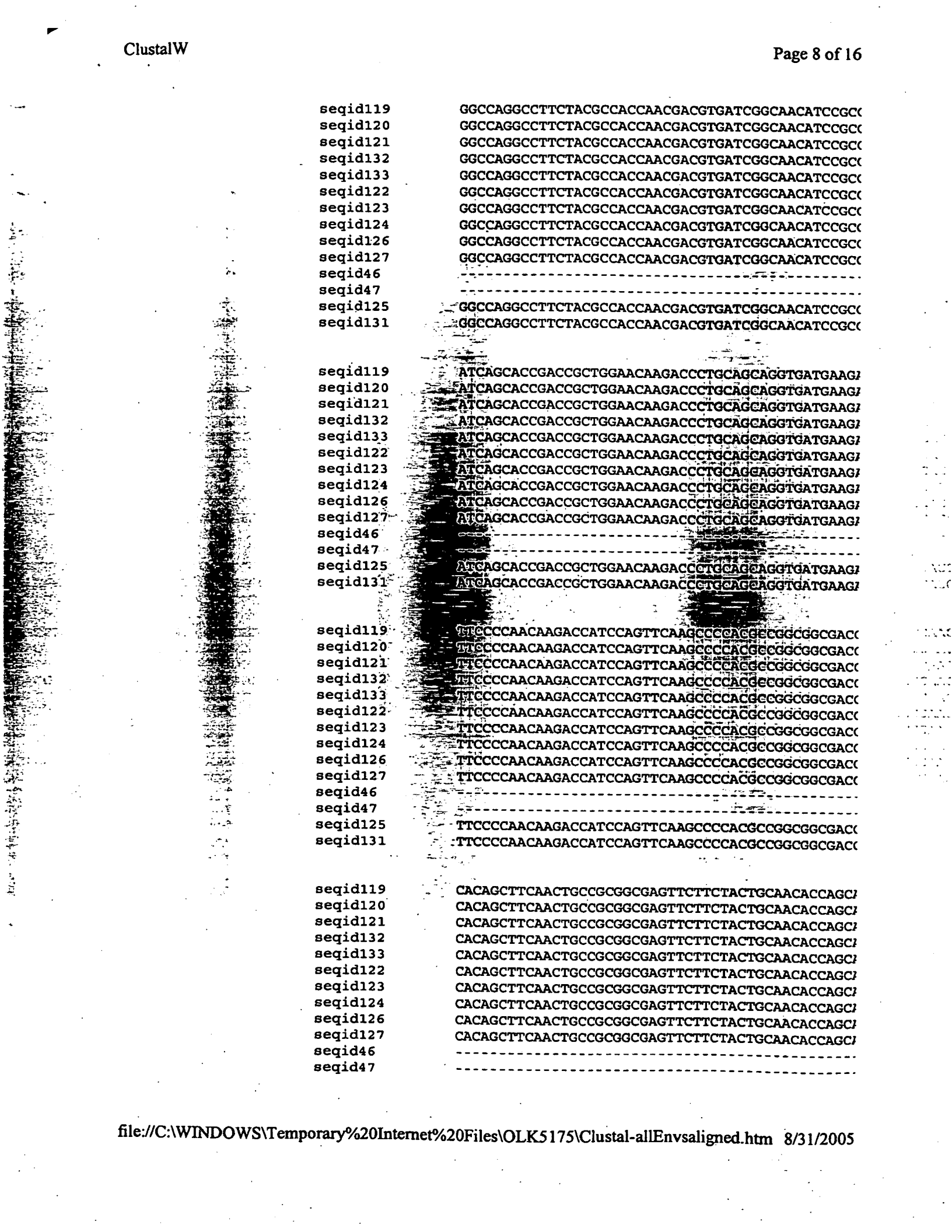
```
CLUSTAL W (1.82) multiple sequence alignment

seqid119        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid120        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid121        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid132        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid133        GAATTCATGAGAGTGATGGGGACACAGAAGAATTGTCAACAATGG
seqid122        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid123        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid124        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid126        GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGG
seqid127        GAATTCATGAGAGTGATGGGGACACAGAAGAATTGTCAACAATGG
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        ---GTCGACGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTC
seqid131        --------------ATGGATGCAATGAAGAGAGGGCTCTGCTGTC
```


seqid119 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid120 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid121 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid132 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid133 TTAGGCTTCTGGATGCTAATGATTTGTAACACCGAGGACCTGTGGC
seqid122 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid123 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid124 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid126 CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGC
seqid127 TTAGGCTTCTGGATGCTAATGATTTGTAACACCGAGGACCTGTGGC
seqid46 ----------
seqid47 ----------
seqid125 GGAGCAGTCTTCGTTTCGCCCAGCGCCAGCACCGAGGACCTGTGGC
seqid131 GGAGCAGTCTTCGTTTCGCCCAGCGCCAGCACCGAGGACCTGTGGC
seqid119 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid120 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid121 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid132 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid133 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid122 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid123 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid124 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid126 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid127 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid46 ----------
seqid47 ----------
seqid125 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid131 GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCA
seqid119 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid120 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid121 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid132 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid133 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid122 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid123 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid124 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid126 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid127 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid46 ----------
seqid47 ----------
seqid125 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid131 TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGC
seqid119 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid120 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid121 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid132 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid133 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid122 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid123 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid124 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid126 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid127 CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGT
seqid46 ----------
seqid47 ----------

```
seqid125        CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATG
seqid131        CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATG


seqid119        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid120        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid121        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid132        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid133        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid122        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid123        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid124        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid126        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid127        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC
seqid131        GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGC


seqid119        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid120        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid121        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid132        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid133        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid122        AAGCTGACCCCCCTGTGCGTG------------------------
seqid123        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid124        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid126        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid127        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC
seqid131        AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAAC


seqid119        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid120        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid121        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid132        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid133        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid122        ----------------------------------------------
seqid123        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid124        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid126        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid127        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC
seqid131        ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATC


seqid119        ATGAAGAACTGCAGCTTCAACGCCGGCGCCGG--------------
seqid120        ATGAAGAACTGCAGCTTCAACGCCGGCGCCGG--------------
seqid121        ATGAAGAACTGCAGCTTCAACGCCGGCGCCGG--------------
seqid132        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAG
seqid133        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAG
seqid122        ------------------------GGCGCCGG--------------
seqid123        ATGAAGAACTGCAGCTTCAACGCCGGCGCCGG--------------
seqid124        ATGAAGAACTGCAGCTTCAACGCCGGCGCCGG--------------
```

```
seqid126        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAGA
seqid127        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAGA
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAGA
seqid131        ATGAAGAACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAGA


seqid119        ----------------------------------------------
seqid120        ----------------------------------------------
seqid121        ----------------------------------------------
seqid132        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC
seqid133        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC
seqid122        ----------------------------------------------
seqid123        ----------------------------------------------
seqid124        ----------------------------------------------
seqid126        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC
seqid127        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC
seqid131        GCCCTGTTCTACCGCCTGGACATCGTGCCCCTGAACGAGAACAGCC


seqid119        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid120        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid121        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid132        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid133        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid122        ---------AACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid123        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid124        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid126        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid127        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
seqid131        CGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA


seqid119        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid120        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid121        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid132        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid133        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid122        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid123        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid124        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid126        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid127        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA
seqid131        CCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGA


seqid119        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAGT
seqid120        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAGT
seqid121        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAGT
seqid132        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAGT
```

```
seqid133        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid122        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid123        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid124        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid126        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid127        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG
seqid131        ACCTTCAACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAG

seqid119        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid120        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid121        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid132        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid133        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid122        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid123        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid124        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid126        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid127        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC
seqid131        AAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCC

seqid119        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid120        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid121        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid132        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid133        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid122        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid123        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid124        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid126        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid127        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
seqid131        ATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC

seqid119        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid120        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid121        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid132        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid133        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid122        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid123        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid124        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid126        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid127        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
seqid131        GTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCC
```

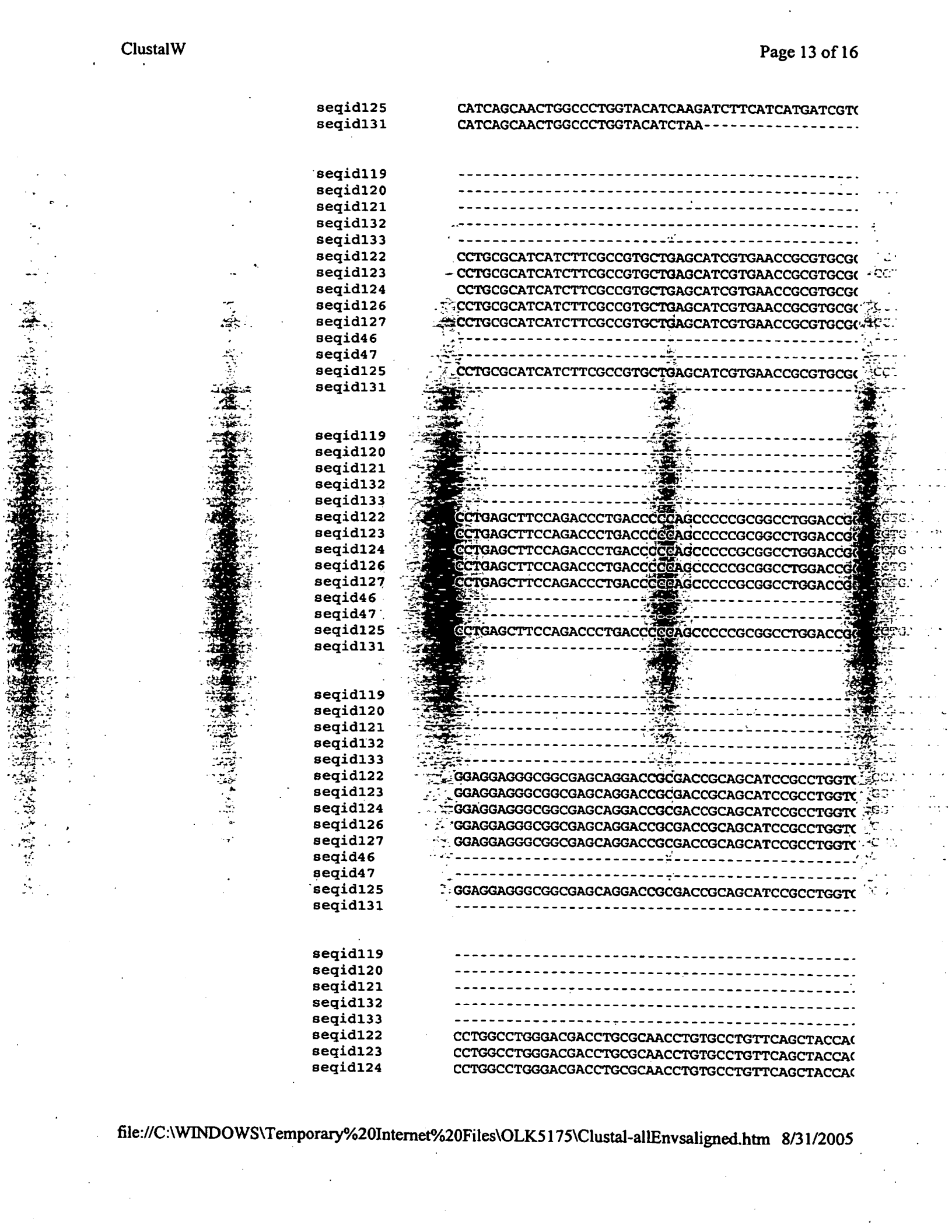

seqid119 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid120 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid121 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid132 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid133 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid122 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid123 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid124 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid126 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid127 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid46 ----------
seqid47 ----------
seqid125 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid131 GGCCAGGCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGC
seqid119 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid120 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid121 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid132 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid133 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid122 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid123 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid124 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid126 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid127 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid46 ----------
seqid47 ----------
seqid125 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid131 ATCAGCACCGACCGCTGGAACAAGACCCTGCAGCAGGTGATGAAG
seqid119 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid120 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid121 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid132 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid133 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid122 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid123 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid124 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid126 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid127 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid46 ----------
seqid47 ----------
seqid125 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid131 TTCCCCAACAAGACCATCCAGTTCAAGCCCCACGCCGGCGGCGAC
seqid119 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid120 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid121 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid132 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid133 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid122 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid123 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid124 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid126 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid127 CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGC
seqid46 ----------
seqid47 ----------

```
seqid125        CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGCA
seqid131        CACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGCA


seqid119        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid120        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid121        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid132        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid133        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid122        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid123        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid124        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid126        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid127        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid46         ------------------------------------------------
seqid47         ------------------------------------------------
seqid125        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA
seqid131        ACCTACCACAGCAACAACGGCACCTACAAGTACAACGGCAACAGCA


seqid119        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid120        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid121        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid132        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid133        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid122        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid123        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid124        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid126        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid127        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid46         CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid47         CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid125        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
seqid131        CTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGC
                ************************************************


seqid119        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid120        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid121        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid132        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid133        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid122        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid123        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid124        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid126        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid127        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid46         GCCCCCCCCATCGCCGGCAACATCACCTGC----------------
seqid47         GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid125        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
seqid131        GCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCC
                *********************************


seqid119        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid120        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid121        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid132        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid133        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid122        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid123        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
seqid124        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGCC
```

```
seqid126        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGC(
seqid127        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGC(
seqid46         ----------------------------------------------
seqid47         CGCGACGGCGGC----------------------------------
seqid125        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGC(
seqid131        CGCGACGGCGGCTTCAACACCACCAACAACACCGAGACCTTCCGC(

seqid119        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid120        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid121        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid132        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid133        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid122        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid123        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid124        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid126        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid127        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(
seqid131        ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTG(

seqid119        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid120        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid121        GGCATCGCCCCCACCAAGGCCATCAGCAGCGTGGTGCAGAGCGAG
seqid132        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid133        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid122        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid123        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid124        GGCATCGCCCCCACCAAGGCCATCAGCAGCGTGGTGCAGAGCGAG
seqid126        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid127        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG
seqid131        GGCATCGCCCCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCGAG

seqid119        ----------------------------------------------
seqid120        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid121        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid132        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid133        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid122        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid123        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid124        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid126        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid127        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(
seqid131        CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCAC(

seqid119        ----------------------------------------------
seqid120        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid121        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid132        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
```

```
seqid133        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid122        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid123        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid124        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid126        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid127        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC
seqid131        CATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTC

seqid119        ----------------------------------------------
seqid120        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid121        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid132        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid133        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid122        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid123        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid124        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid126        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid127        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC
seqid131        CCTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACC

seqid119        ----------------------------------------------
seqid120        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid121        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid132        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid133        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid122        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid123        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid124        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid126        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid127        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC
seqid131        GCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGAC

seqid119        ----------------------------------------------
seqid120        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid121        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid132        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid133        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid122        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid123        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid124        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid126        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid127        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
seqid131        CATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCGTGCCC
```

```
seqid119        ----------------------------------------
seqid120        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid121        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid132        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid133        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid122        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid123        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid124        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid126        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid127        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC
seqid131        GAGCAACAAGAGCGAGAAGGACATCTGGGACAACATGACCTGGATC

seqid119        ----------------------------------------
seqid120        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid121        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid132        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid133        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid122        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid123        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid124        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid126        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid127        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC
seqid131        GATCAGCAACTACACCGGCCTGATCTACAACCTGCTGGAGGACAGC

seqid119        ----------------------------------------
seqid120        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid121        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid132        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid133        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid122        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid123        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid124        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid126        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid127        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
seqid125        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC
seqid131        GAAGAACGAGAAGGACCTGCTGGAGCTGGACAAGTGGAACAACCTC

seqid119        ----------------------------------------
seqid120        CATCAGCAACTGGCCCTGGTACATCTAACTCGAG----------
seqid121        CATCAGCAACTGGCCCTGGTACATCTAACTCGAG----------
seqid132        CATCAGCAACTGGCCCTGGTACATCTAACTCGAG----------
seqid133        CATCAGCAACTGGCCCTGGTACATCTAACTCGAG----------
seqid122        CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGTC
seqid123        CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGTC
seqid124        CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGTC
seqid126        CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGTC
seqid127        CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGTC
seqid46         ----------------------------------------
seqid47         ----------------------------------------
```


seqid125 CATCAGCAACTGGCCCTGGTACATCAAGATCTTCATCATGATCGT
seqid131 CATCAGCAACTGGCCCTGGTACATCTAA
seqid119
seqid120
seqid121
seqid132
seqid133
seqid122 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid123 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid124 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid126 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid127 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid46
seqid47
seqid125 CCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCG
seqid131
seqid119
seqid120
seqid121
seqid132
seqid133
seqid122 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid123 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid124 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid126 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid127 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid46
seqid47
seqid125 CCTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCG
seqid131
seqid119
seqid120
seqid121
seqid132
seqid133
seqid122 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid123 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid124 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid126 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid127 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid46
seqid47
seqid125 GGAGGAGGGCGGCGAGCAGGACCGCGACCGCAGCATCCGCCTGGTC
seqid131
seqid119
seqid120
seqid121
seqid132
seqid133
seqid122 CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCA
seqid123 CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCA
seqid124 CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCA

```
seqid126        CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCAC
seqid127        CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCAC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CCTGGCCTGGGACGACCTGCGCAACCTGTGCCTGTTCAGCTACCAC
seqid131        ----------------------------------------------

seqid119        ----------------------------------------------
seqid120        ----------------------------------------------
seqid121        ----------------------------------------------
seqid132        ----------------------------------------------
seqid133        ----------------------------------------------
seqid122        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid123        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid124        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid126        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid127        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        CATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGC
seqid131        ----------------------------------------------

seqid119        ----------------------------------------------
seqid120        ----------------------------------------------
seqid121        ----------------------------------------------
seqid132        ----------------------------------------------
seqid133        ----------------------------------------------
seqid122        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid123        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid124        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid126        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid127        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GCGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTAC
seqid131        ----------------------------------------------

seqid119        ----------------------------------------------
seqid120        ----------------------------------------------
seqid121        ----------------------------------------------
seqid132        ----------------------------------------------
seqid133        ----------------------------------------------
seqid122        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid123        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid124        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid126        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid127        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid46         ----------------------------------------------
seqid47         ----------------------------------------------
seqid125        GAAGAAGAGCGCCATCAGCCTGCTGGACACCATCGCCATCACCGTC
seqid131        ----------------------------------------------

seqid119        ----------------------------------------------
seqid120        ----------------------------------------------
seqid121        ----------------------------------------------
seqid132        ----------------------------------------------
```

```
seqid133        ----------------------------------------------------
seqid122        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid123        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid124        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid126        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid127        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid46         ----------------------------------------------------
seqid47         ----------------------------------------------------
seqid125        CCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCCTGAAC
seqid131        ----------------------------------------------------


seqid119        ---------------------------------------- 1
seqid120        ---------------------------------------- 2
seqid121        ---------------------------------------- 2
seqid132        ---------------------------------------- 3
seqid133        ---------------------------------------- 3
seqid122        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2397
seqid123        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2529
seqid124        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2529
seqid126        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2616
seqid127        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2616
seqid46         ---------------------------------------- 16
seqid47         ---------------------------------------- 17
seqid125        CCGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG 2613
seqid131        ---------------------------------------- 18
```

PLEASE NOTE: Showing colors on large alignments is slow.

Show Colors | View Alignment File

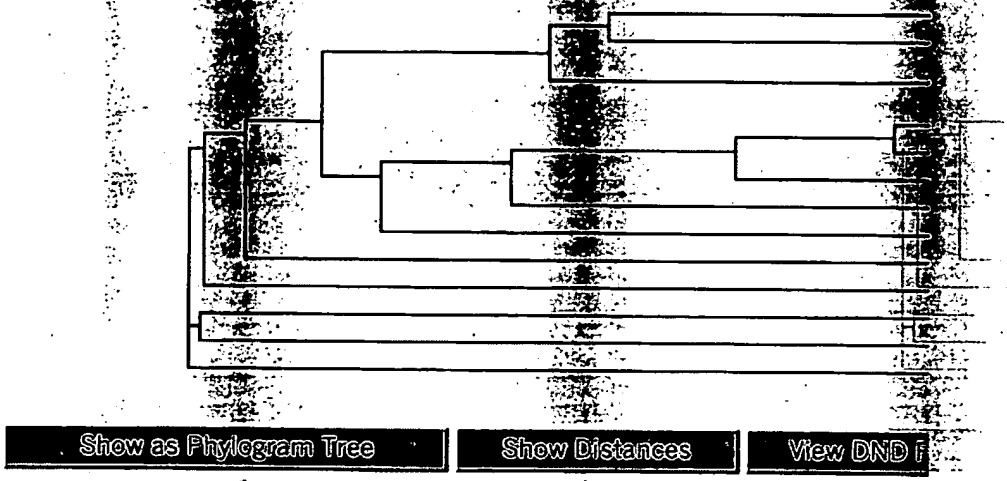
Guide Tree:

Show as Phylogram Tree | Show Distances | View DND F

```
(
(
(
(
(
(
seqid46:-0.00705,
seqid47:0.00705)
:0.00360,
(
seqid125:-0.00719,
seqid131:0.00817)
:0.02358)
:0.01417,
(
(
(
(
```

```
seqid122:0.01713,
seqid123:-0.01713)
:0.00978,
seqid124:-0.00734)
:0.01374,
seqid126:-0.01249)
:0.00803,
seqid127:-0.00255)
:0.00380)
:0.00475,
seqid133:0.00545)
:0.00253,
seqid132:0.00078)
:0.00102,
(
seqid119:0.00677,
seqid120:-0.00066)
:0.00081,
seqid121:0.00274);
```

Cladogram



Right-click on the above tree to see display options.
Problems printing? Read how to print a Phylogram or Cladogram.

```
CLUSTAL W (1.82) multiple sequence alignment


seqidno120      GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGGTGGATCTGGGGCATC 60
seqidno121      GAATTCATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGGTGGATCTGGGGCATC 60
                ************************************************************

seqidno120      CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGGTGACCGTGTACTAC 120
seqidno121      CTGGGCTTCTGGATGCTGATGATCTGCAACACCGAGGACCTGTGGGTGACCGTGTACTAC 120
                ************************************************************

seqidno120      GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCC 180
seqidno121      GGCGTGCCCGTGTGGCGCGACGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCC 180
                ************************************************************

seqidno120      TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACCGACCCCAAC 240
seqidno121      TACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACCGACCCCAAC 240
                ************************************************************

seqidno120      CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATG 300
seqidno121      CCCCAGGAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATG 300
                ************************************************************

seqidno120      GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG 360
seqidno121      GCCGACCAGATGCACGAGGACGTGATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG 360
                ************************************************************

seqidno120      AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAACGTGACCGGCAACCGC 420
seqidno121      AAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAACGTGACCGGCAACCGC 420
                ************************************************************

seqidno120      ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATCTACAACATCGAGGAG 480
seqidno121      ACCGTGACCGGCAACAGCACCAACAACACCAACGGCACCGGCATCTACAACATCGAGGAG 480
                ************************************************************

seqidno120      ATGAAGAACTGCAGCTTCAACGCCGGCGCCGGCCGCCTGATCAACTGCAACACCAGCACC 540
seqidno121      ATGAAGAACTGCAGCTTCAACGCCGGCGCCGGCCGCCTGATCAACTGCAACACCAGCACC 540
                ************************************************************

seqidno120      ATCACCCAGGCCTGCCCCAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCC 600
seqidno121      ATCACCCAGGCCTGCCCCAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCC 600
                ************************************************************

seqidno120      GCCGGCTACGCCATCCTGAAGTGCAACAACAAGACCTTCAACGGCACCGGCCCCTGCTAC 660
seqidno121      GCCGGCTACGCCATCCTGAAGTGCAACAACAAGACCTTCAACGGCACCGGCCCCTGCTAC 660
                ************************************************************

seqidno120      AACGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGAGCACCCAGCTGCTG 720
seqidno121      AACGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGAGCACCCAGCTGCTG 720
                ************************************************************

seqidno120      CTGAACGGCAGCCTGGCCGAGGAGGGCATCATCATCCGCAGCGAGAACCTGACCGAGAAC 780
seqidno121      CTGAACGGCAGCCTGGCCGAGGAGGGCATCATCATCCGCAGCGAGAACCTGACCGAGAAC 780
                ************************************************************

seqidno120      ACCAAGACCATCATCGTGCACCTGAACGAGAGCGTGGAGATCAACTGCACCCGCCCCAAC 840
seqidno121      ACCAAGACCATCATCGTGCACCTGAACGAGAGCGTGGAGATCAACTGCACCCGCCCCAAC 840
                ************************************************************

seqidno120      AACAACACCCGCAAGAGCGTGCGCATCGGCCCCGGCCAGGCCTTCTACGCCACCAACGAC 900
```

```
seqidno121      AACAACACCCGCAAGAGCGTGCGCATCGGCCCCGGCCAGGCCTTCTACGCCACCAACGAC 900
                ************************************************************

seqidno120      GTGATCGGCAACATCCGCCAGGCCCACTGCAACATCAGCACCGACCGCTGGAACAAGACC 960
seqidno121      GTGATCGGCAACATCCGCCAGGCCCACTGCAACATCAGCACCGACCGCTGGAACAAGACC 960
                ************************************************************

seqidno120      CTGCAGCAGGTGATGAAGAAGCTGGGCGAGCACTTCCCCAACAAGACCATCCAGTTCAAG 1020
seqidno121      CTGCAGCAGGTGATGAAGAAGCTGGGCGAGCACTTCCCCAACAAGACCATCCAGTTCAAG 1020
                ************************************************************

seqidno120      CCCCACGCCGGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCCGCGGCGAGTTC 1080
seqidno121      CCCCACGCCGGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCCGCGGCGAGTTC 1080
                ************************************************************

seqidno120      TTCTACTGCAACACCAGCAACCTGTTCAACAGCACGTACCACAGCAACAACGGCACCTAC 1140
seqidno121      TTCTACTGCAACACCAGCAACCTGTTCAACAGCACCTACCACAGCAACAACGGCACCTAC 1140
                ************************************************************

seqidno120      AAGTACAACGGCAACAGCAGCAGCCCCATCACCCTGCAGTGCAAGATCAAGCAGATCGTG 1200
seqidno121      AAGTACAACGGCAACAGCAGCAGCCCCATCACCCTGCAGTGCAAGATCAAGCAGATCGTG 1200
                ************************************************************

seqidno120      CGCATGTGGCAGGGCGTGGGCCAGGCCACCTACGCCCCCCCCATCGCCGGCAACATCACC 1260
seqidno121      CGCATGTGGCAGGGCGTGGGCCAGGCCACCTACGCCCCCCCCATCGCCGGCAACATCACC 1260
                ************************************************************

seqidno120      TGCCGCAGCAACATCACCGGCATCCTGCTGACCCGCGACGGCGGCTTCAACACCACCAAC 1320
seqidno121      TGCCGCAGCAACATCACCGGCATCCTGCTGACCCGCGACGGCGGCTTCAACACCACCAAC 1320
                ************************************************************

seqidno120      AACACCGAGACCTTCCGCCCCGGCGGCGGCGACATGCGCGACAACTGGCGCAGCGAGCTG 1380
seqidno121      AACACCGAGACCTTCCGCCCCGGCGGCGGCGACATGCGCGACAACTGGCGCAGCGAGCTG 1380
                ************************************************************

seqidno120      TACAAGTACAAGGTGGTGGAGATCAAGCCCCTGGGCATCGCCCCCACCAAGGCCAAGCGC 1440
seqidno121      TACAAGTACAAGGTGGTGGAGATCAAGCCCCTGGGCATCGCCCCCACCAAGGCCATCAGC 1440
                ************************************************************

seqidno120      CGCGTGGTGCAGCGCGAGAAGCGCGCCGTGGGCATCGGCGCCGTGTTCCTGGGCTTCCTG 1500
seqidno121      AGCGTGGTGCAGAGCGAGAAGAGCGCCGTGGGCATCGGCGCCGTGTTCCTGGGCTTCCTG 1500
                ************ ******** ****************************************

seqidno120      GGCGCCGCCGGCAGCACCATGGGCGCCGCCAGCATCACCCTGACCGTGCAGGCCCGCCAG 1560
seqidno121      GGCGCCGCCGGCAGCACCATGGGCGCCGCCAGCATCACCCTGACCGTGCAGGCCCGCCAG 1560
                ************************************************************

seqidno120      CTGCTGAGCGGCATCGTGCAGCAGCAGAGCAACCTGCTGAAGGCCATCGAGGCCCAGCAG 1620
seqidno121      CTGCTGAGCGGCATCGTGCAGCAGCAGAGCAACCTGCTGAAGGCCATCGAGGCCCAGCAG 1620
                ************************************************************

seqidno120      CACATGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGCTGGCCATC 1680
seqidno121      CACATGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGCTGGCCATC 1680
                ************************************************************

seqidno120      GAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCCGCCTGATC 1740
seqidno121      GAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCCGCCTGATC 1740
                ************************************************************

seqidno120      TGCACCACCGCCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCGAGAAGGACATCTGG 1800
```

```
seqidno121      TGCACCACCGCCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCGAGAAGGACATCTGG 1800
                ************************************************************

seqidno120      GACAACATGACCTGGATGCAGTGGGACCGCGAGATCAGCAACTACACCGGCCTGATCTAC 1860
seqidno121      GACAACATGACCTGGATGCAGTGGGACCGCGAGATCAGCAACTACACCGGCCTGATCTAC 1860
                ************************************************************

seqidno120      AACCTGCTGGAGGACAGCCAGAACCAGCAGGAGAAGAACGAGAAGGACCTGCTGGAGCTG 1920
seqidno121      AACCTGCTGGAGGACAGCCAGAACCAGCAGGAGAAGAACGAGAAGGACCTGCTGGAGCTG 1920
                ************************************************************

seqidno120      GACAAGTGGAACAACCTGTGGAACTGGTTCGACATCAGCAACTGGCCCTGGTACATCTAA 1980
seqidno121      GACAAGTGGAACAACCTGTGGAACTGGTTCGACATCAGCAACTGGCCCTGGTACATCTAA 1980
                ************************************************************

seqidno120      CTCGAG 1986
seqidno121      CTCGAG 1986
                ******
```

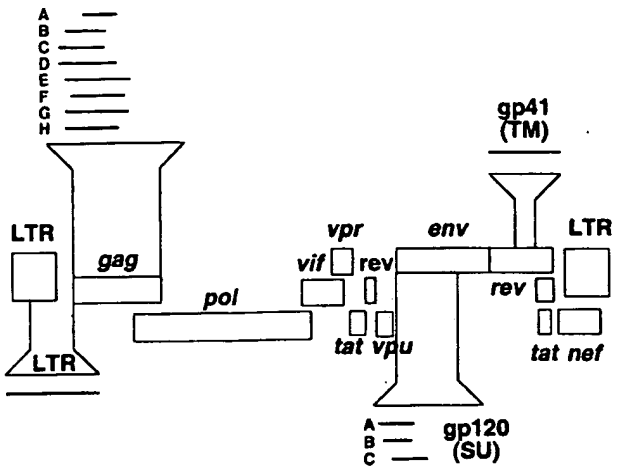


FIG. 1. HIV-1 subgenomic regions utilized for epidemiological linkage analysis. A schematic representation of the HIV-1 genome is shown, with brackets denoting the subgenomic *gag*, gp120, gp41, and LTR fragments that were amplified for diagnostic PCR analysis. Overlapping *gag* and gp120 regions are denoted by capital letters (*gagA* to *-H*; gp120A to -C) and are referred to in Tables 2 to 4.

the ability to confirm, with a high level of confidence, epidemiological linkage of HIV-1 transmission between members of all putative transmission pairs.

Molecular analyses of suspected transmission links have been widely used to characterize localized HIV-1 outbreaks, mother-to-infant transmission, sexual transmission, sharing of contaminated needles, donation of contaminated blood, receipt of contaminated clotting reagent, nosocomial transmissions from health care workers, and intrafamilial contacts (1, 4, 5, 6, 13, 18, 21, 29, 36, 42, 44). In all of these cases, the establishment of epidemiological linkage relied on the documentation of closer genetic relatedness between viruses infecting the suspected transmission pair(s) compared to control viruses isolated from unrelated individuals in the same region. Here we developed a similar approach to confirm (or refute) heterosexual transmission among discordant couples within the ZUHRP cohort.

Blood samples were collected between 1996 and 2000 from both partners of 149 (of a total of 162) discordant couples in whom seroconversion had been documented. The time period between the last negative and the first positive blood tests for the seroconvertor (which in most cases was also the blood sample used for linkage analysis) was 4.9 months on average, but in some cases it extended up to 4 years. High-molecular-weight DNA was extracted from whole blood or Ficoll gradient-purified peripheral blood mononuclear cells by using the QIAamp Blood Kit (Qiagen, Valencia, Calif.). For a small number of couples, DNA was extracted from dried blood spots (9). Because of the known variability of HIV-1, different regions of the HIV-1 genome were targeted for PCR amplification, resulting in comparisons of *gag*, gp120, gp41, and/or long terminal repeat (LTR) regions as shown in Fig. 1. Although the gp41 primers were by far the most cross-reactive, the suitability of this primer set was discovered only after alternative genomic regions from a number of transmission pairs had already been analyzed (43). LTR, gp120, and gp41 primers and amplification conditions have been described previously (15, 16). The primers that were used to amplify sequences within *gag* were cgagA 5′-TGATAAAACCTCCAATTCCCCCTA T-3′ and PBS1A 5′-TTTGCCTGTACTGGGTCTCTCTGGT T-3′ in the first round and cgagB 5′-AATACTGTATCATCT GCTCCTGTATC-3′ and PBS1B 5′-GCTTAAGCCTCAATA AAGCTTGCCTT-3′ in the second round. PCR products were sequenced directly, using cycle sequencing and dye terminator methodologies, on an automated DNA sequencer (model 377A; Applied Biosystems, Inc., Foster City, Calif.). Both strands of the PCR products were sequenced (sequences are available under GenBank accession numbers AF404868 through AF405203, AF406742, and AF406743). Although population-based sequencing was used to allow analysis of the predominant viral form in each individual, the number of ambiguous base pairs in the entire data set was <0.3%.

To establish suitable linkage criteria for HIV-1 strains infecting the Zambian couples, amplified viral sequences were first subjected to preliminary phylogenetic tree analyses to identify all circulating HIV-1 group M subtypes (not shown). Full-length and nonrecombinant reference sequences representing these subtypes were then obtained from the Los Alamos HIV Sequence Database (Table 1) and subjected to pairwise sequence comparisons in the genomic regions corre-

TABLE 1. Subtype-specific full-length reference sequences from the HIV Sequence Database

Subtype_sequence name	Accession no.	Reference
A_Q2317	AF004885	31
A_SE8891	AF069673	7
A_SE8538	AF069669	7
A_SE6594	AF069672	7
A_SE7535	AF069671	7
A_SE7253	AF069670	7
A_SE8131	AF107771	7
A_U455	M62320	28
A_92UG037.1	U51190	14
C_96BW04.07	AF110963	26
C_96BW11B01	AF110971	26
C_96BW15C02	AF110974	26
C_96BW05.04	AF110968	26
C_96BW16.26	AF110978	26
C_96BW12.10	AF110972	26
C_96BW17A09	AF110979	26
C_96BW01B21	AF110960	26
C_C2220	U46016	35
C_94IN11246	AF067159	24
C_98BR004	AF286228	33
C_98IS002.5	AF286233	33
C_98TZ013.10	AF286234	33
C_98TZ017.2	AF286235	33
C_97ZA012.1	AF286227	33
D_94UG114.1	U88824	14
D_84ZR085	U88822	14
D_NDK	M27323	37
D_ELI	K03454	2
D_Z2Z6	M22639	38
G_DRCBL	AF084936	27
G_HH8793	AF061641	8
G_92NG083.2	U88826	14
G_SE6165	AF061642	8
J_SE91733	AF082395	23
J_SE92809	AF082394	23

JOURNAL OF VIROLOGY, May 1999, p. 4427–4432
0022-538X/99/$04.00+0

Vol. 73, No. 5

NOTES

Molecular Cloning and Phylogenetic Analysis of Human Immunodeficiency Virus Type 1 Subtype C: a Set of 23 Full-Length Clones from Botswana

VLADIMIR A. NOVITSKY,[1] MONTY A. MONTANO,[1] MARY F. McLANE,[1] BORIS RENJIFO,[1] FREDRIK VANNBERG,[1] BRIAN T. FOLEY,[2] THUMBI P. NDUNG'U,[1] MAFIZUR RAHMAN,[3] MOEKETSI J. MAKHEMA,[4] RICHARD MARLINK,[1] AND MAX ESSEX[1]*

Harvard AIDS Institute, Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115,[1] and Theoretical Biology and Biophysics, Group T-10, Los Alamos National Laboratory, Los Alamos, New Mexico 87545,[2] and AIDS/STD Unit[3] and Princess Marina Hospital,[4] Gaborone, Botswana

Received 16 October 1998/Accepted 4 February 1999

To better understand the virological aspect of the expanding AIDS epidemic in southern Africa, a set of 23 near-full-length clones of human immunodeficiency virus type 1 (HIV-1) representing eight AIDS patients from Botswana were sequenced and analyzed phylogenetically. All study viruses from Botswana belonged to HIV-1 subtype C. The interpatient diversity of the clones from Botswana was higher than among full-length isolates of subtype B or among a set of full-length HIV-1 genomes of subtype C from India (mean value of 9.1% versus 6.5 and 4.3%, respectively; $P < 0.0001$ for both comparisons). Similar results were observed in all genes across the entire viral genome. We suggest that the high level of HIV-1 diversity might be a typical feature of the subtype C epidemic in southern Africa. The reason or reasons for this diversity are unclear, but may include an altered replication efficiency of HIV-1 subtype C and/or the multiple introduction of different subtype C viruses.

The majority of new human immunodeficiency virus (HIV) infections in the global AIDS epidemic are appearing in sub-Saharan Africa and Southeast Asia. Compared with the situation a decade ago, the main AIDS epidemics have shifted from central and eastern Africa to the southern regions. The most severe HIV epidemics have recently afflicted such southern African countries as Zimbabwe, Zambia, Namibia, South Africa, and Botswana (43). HIV-1 subtype C has been estimated to account for 48% of HIV-1 infections worldwide and 51.5% of HIV-1 infections in Africa (4, 7, 14–16, 21, 31), where the main mode of transmission is heterosexual (43, 44, 47).

A rapid expansion of the HIV-1 epidemic in Botswana has occurred since the early to mid 1990s. According to the UNAIDS and World Health Organization (WHO) Global HIV/AIDS & STD Surveillance data, HIV prevalence among antenatal clinic attendees tested in the major urban areas of Botswana (Gaborone, Francistown, and Selebi-Phikwe) increased from 6% in 1990 to 39% in 1997 (range of 34 to 43%) (42). Among women 20 to 29 years of age, 43 to 44% tested HIV positive. Outside of the major urban areas, median HIV prevalence increased from no evidence of infection in 1985 to 1987 to 34% in 1997. In 1997, HIV prevalence in Botswana ranged from 28 to 38%. As such, locally circulating HIV-1 needs to be characterized thoroughly, and vital information about the nature of the epidemic should be extended (2, 4, 7, 21, 31, 37, 45–47). Moreover, Botswana's central geographic position makes a comprehensive HIV-1 molecular epidemiological study that much more urgent, because it may serve as an example of the burgeoning epidemic in southern Africa.

In this study, we report the molecular cloning and phylogenetic analysis of 23 near-full-length clones from Botswana. All of them were identified as belonging to HIV-1 subtype C and demonstrated high levels of intersample diversity across the entire viral genome. By providing new genetic information regarding locally circulating viruses, this study may contribute to AIDS vaccine design for the southern Africa region countries and, in particular, for Botswana.

Specimens for this study were selected from HIV-seropositive patients in Gaborone, Botswana. All HIV-1 infections in this study were likely to be heterosexually acquired. The times of infection were not known. The HIV-1-seropositive status of patients was confirmed by enzyme-linked immunosorbent assay and Western blot analysis. Clinical classification was performed by using the 1987 Centers for Disease Control and Prevention (CDC) revised classification (9) (data not shown).

Genomic DNA was obtained directly from the patients' peripheral blood mononuclear cells (PBMCs)—buffy coats—without passage through cell culture or donor PBMCs. All clones in this study were amplified in heminested PCR with three primers from the LA set (18) or their modifications. The Expand Long Template PCR system (Boehringer Mannheim, Indianapolis, Ind.) was used according to the manufacturer's instructions. Gel purification of the first-round PCR product was essential for direct amplification of 9.0-kb fragments from uncultured PBMCs. Estimation of the expanded PCR sensitivity (based on 8E5/LAV) revealed a successful amplification of the 9.0-kb fragment in the first round when at least 8×10^2 to

* Corresponding author. Mailing address: Department of Immunology and Infectious Diseases, Harvard School of Public Health, FXB-402, 651 Huntington Ave., Boston, MA 02115. Phone: (617) 432-0975. Fax: (617) 739-8348. E-mail: messex@sph.harvard.edu.

DISCLOSES AF110968, AF110975

ISOLATE C-96BW05.04

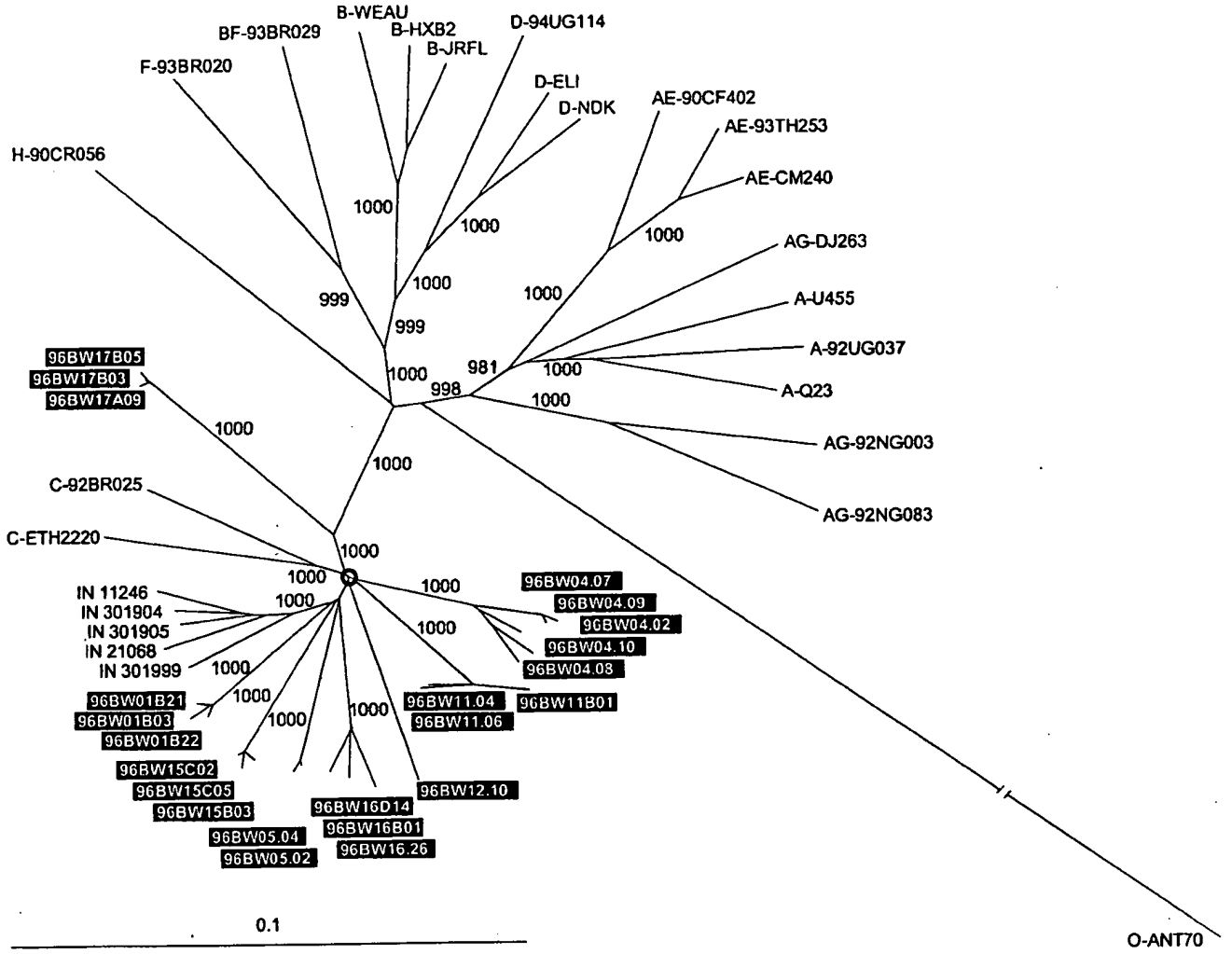


FIG. 1. Phylogenetic relationship of the newly characterized full-length clones from Botswana (boxed in black) to other representative full-length HIV-1 sequences of subtypes A, B, C, D, F, and H and recombinant subtypes AE, AG, and BF. Full-length subtype C sequences from India were also included in the analysis. A neighbor-joining tree was constructed on the basis of the hidden Markov model nucleotide alignment of full-length HIV-1 genomes. Subtype O ANT70 sequence was used as an outgroup. Values along the branches indicate the bootstrap values that support branching (out of a 1,000 resampling).

4×10^3 proviral copies were present in the reaction (data not shown). These results were consistent with those from other studies (10, 39). The TA pCR2.1 TOPO system (Invitrogen, Carlsbad, Calif.) and JM109 competent cells (Promega Corporation, Madison, Wis.) were used for cloning. Positive colonies were screened by PCR. To obtain sufficient plasmids for sequence analysis, we amplified the constructs under the previously described conditions with some modifications (41). Purified plasmid DNA served as a template for sequencing. Both-strand sequencing was combined with a strategy involving overlapping sequences. Dye terminator sequencing on an automated DNA Sequenator (model 373A; Applied Biosystems, Inc., Foster City, Calif.) was used.

A multiple alignment procedure for the full-length HIV genome was performed by using the hidden Markov model. Constructed through the HIV-1 HMMER computer program of the Los Alamos National Laboratory, the model has been previously shown to provide the best description of the true nucleotide substitution pattern of HIV-1 *gag* and *env* genes (26). The HIV-1 HMMER model (11, 12) constructed at Los Alamos National Laboratory for the full-length HIV-1 genomes (24) was employed. Sixty full-length reference sequences were included in the alignment from the GenBank data bank (5). The 3′ end of the alignment, which included the *nef* coding region and 3′ long terminal repeat (LTR), was adjusted manually. The pairwise evolutionary distances from nucleotide sequences were computed by the DNADIST program under Kimura's two-parameter model (17). All alignments were globally gap stripped for the generation of the trees. The transition/transversion ratio parameters were set at 3.0 for the *gag* gene, 1.5 for the *env* gene, 1.42 for the V1-V2 and V3 fragments, and 2.0 for the other viral loci (25). A tree was drawn by the Njplot (33) and TreeView (32) programs. To analyze patterns of variability along the HIV-1 genomes, the program SWAN, which utilizes a "sliding window" approach was used (34). Positions with gaps either were or were not excluded from the analysis. The variability distribution was estimated as an entropy function of the nucleotide variation observed at a particular position. The Recombinant Identification Program (RIP) (40) and HIV-1 Subtyping Basic BLAST (3) were used in searching for recombination among the clones studied.

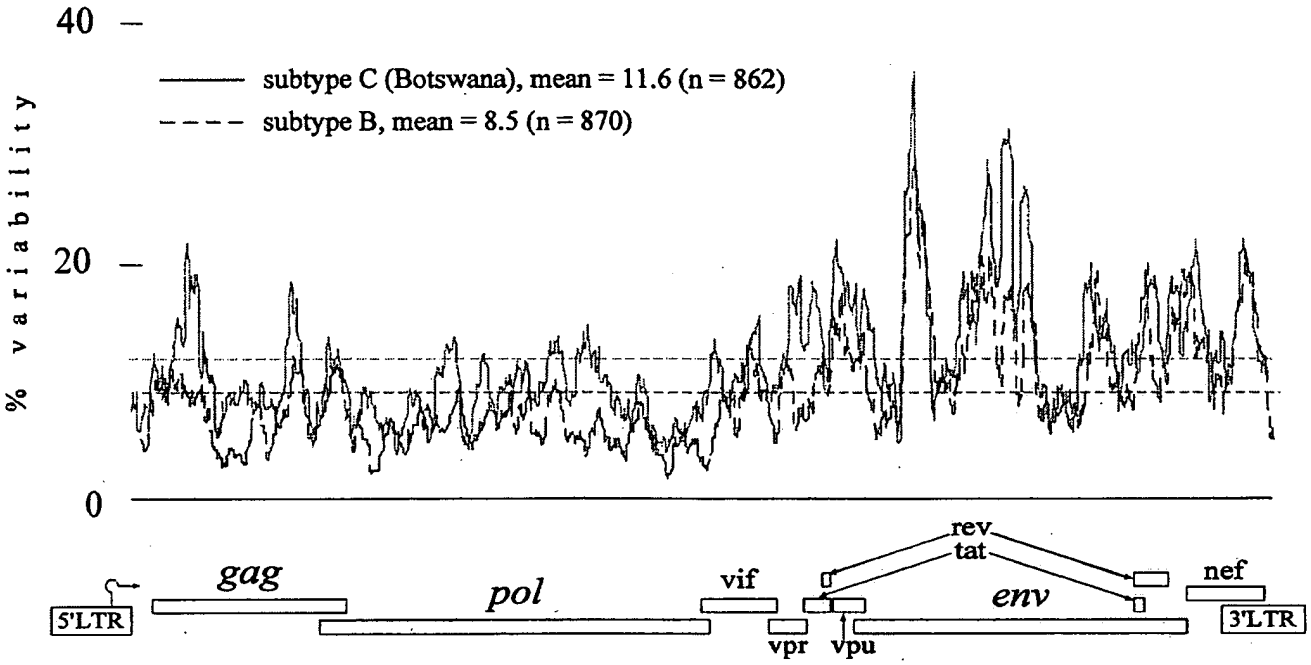


FIG. 2. Variability plots comparing subtype B and C (Botswana) sequences across the entire HIV-1 genome. The variability distribution was estimated as an entropy function of the nucleotide variation in the SWAN program based on the hidden Markov model alignment of the complete HIV-1 genome. The subtype B sequences used in the analysis were DH123, 89.6, RF, WEAU, OYI, HXB2, JRFL, and YU2. *n*, number of sliding window sites across the HIV-1 genome with gap stripping.

Sequence analysis of the Botswana HIV-1 revealed that 10 of 23 clones had an intact genomic organization with open reading frames. The other clones had point mutations and/or insertions and deletions resulting in frameshifts, disabled start codons, or premature stop codons. No major deletions or rearrangements were observed. Determined length polymorphism among studied sequences was limited to the *vpu* (15- to 18-nucleotide [nt] insertion at the 5′ end), *env* (from 3- to 9-nt deletions to 48-nt insertions, GIGRGQ motif in the BW17 V3 loop), 2nd exon of *rev* (a 13-amino-acid truncation at the 3′ end), *nef* (6- to 15-nt deletions in some clones), and regulatory regions of the LTR (three or four NF-κB sites with GGGAC TTTCT as a potential fourth NF-κB in two clones of BW05).

An evolutionary tree in Fig. 1 shows the phylogenetic relationship of the full-length Botswana clones to other representative full-length HIV-1 sequences. All Botswana sequences, a set from India (27), and two subtype C reference sequences (C-ETH2220 [38] and C-92BR025 [19]) clustered together, forming a compelling subtype C outcropping on the phylogenetic tree. This cluster was separated from other HIV-1 sequences by the extremely high bootstrap value of 1,000 (out of 1,000 resampling). Phylogenetic relationships within the subtype C bush were also noteworthy. Assuming that the circled node at the center of the bush could represent the potential ancestral subtype C node, we observed the following. (i) The star-like phylogeny of the subtype C bush together with its branching order may demonstrate the relatively high diversity of the Botswana samples. (ii) Four Botswana sample clones (BW01, BW05, BW15, and BW16), together with all five sequences from India, formed a potential subcluster, although the bootstrap value was not high. (iii) All Indian samples were separated by bootstrap values of 1,000, possibly reflecting a "founder effect" among these sequences. (iv) Three Botswana sample clones (BW04, BW11, and BW12) may represent individualized groups of sequences inherited from a common subtype C ancestor. (v) Two reference sequences ETH2220 and 92BR025 deviated together with a high bootstrap value (1,000), possibly reflecting another subtype C subcluster differing significantly from Botswana or Indian samples. (vi) One of the Botswana samples (BW17) strayed rather far off the main subtype C bush and may be the least representative of Botswana HIV-1 samples. (vii) The topology of the Botswana clones confirms that clones of the same samples are closely related to each other based on full-length genome sequences. A multilocus analysis was congruent with full-genome phylogenetic analysis and confirmed clustering of newly derived Botswana clones within subtype C across the entire HIV-1 genome (data not shown).

To characterize the level of variability among Botswana clones across the entire HIV-1 genome, we performed SWAN program analysis as an entropy function of the nucleotide variation. The Botswana set had greater variability than subtype B samples (Fig. 2) (mean values of 11.6 and 8.5%, respectively, for gap-stripped analysis). The profiles of viral variability across the HIV-1 genome were similar among subtype B and C viruses. Comparison of gap-stripped and gap-nonstripped plots revealed that the differences in mean values between the two methods of computing and the shape of variability plot profiles were not significant (data not shown). Gap stripping slightly decreased the mean value and the number of sliding window sites across the genome. It also hid the extreme regions with the highest level of variability. Both variability when measured as an entropy function in this study and when described before diversity as a pairwise comparison of the sequence (19) exhibited similar profiles of variable and conservative genomic regions. Variability plots (especially non-gap stripped) revealed higher peaks in variable regions than diversity plots.

Table 1 depicts the high degree of intersample diversity across the entire HIV-1 genome among Botswana clones compared with subtype B and C sequences from India (27). Be-

TABLE 1. Intersample HIV-1 diversity in this study[a]

Gene or region	Mean % (range) HIV-1 diversity[b]			
	Subtype C, Botswana	Subtype B		Subtype C, India
		AIDS patients (8 sequences)	23 isolates	
gag	7.9 (5.9–9.2)	5.6* (3.6–9.6)	4.9* (1.0–8.6)	3.4* (2.4–4.6)
pol	5.9 (4.3–7.8)	4.1* (2.9–5.4)	3.9* (0.8–6.2)	2.6* (1.5–3.7)
vif	7.5 (4.3–12.8)	6.3† (3.7–9.3)	6.4‡ (0–9.5)	3.2* (1.8–4.3)
vpr	9.8 (5.3–14.9)	6.2* (4.3–8.4)	5.9* (1.7–10.8)	4.8* (2.1–8.1)
tat	9.9 (6.9–14.9)	7.5* (4.6–10.7)	7.1* (3.4–10.7)	3.7* (2.7–5.1)
rev	10.4 (7.7–16.6)	9.4§ (4.2–14.2)	8.2* (3.3–14.2)	4.7* (2.8–5.7)
vpu	13.9 (10.1–18.8)	10.8* (6.1–14.3)	9.5* (4.2–15.8)	5.1* (2.1–8.1)
env	12.4 (10.4–14.5)	9.5* (6.2–11.9)	9.5* (5.9–12.7)	6.9* (5.1–9.3)
V1-V2	25.7 (15.9–36.7)	18.0* (8.6–26.1)	19.0* (8.6–30.8)	25.8 (14.2–34.8)
V3	14.4 (11.5–18.9)	11.6* (6.7–17.3)	12.1* (6.0–18.2)	6.4* (5.1–7.9)
nef	11.3 (8.0–15.1)	9.8‖ (6.3–15.7)	9.2* (3.5–16.3)	5.2* (4.4–6.4)
3′ LTR	9.7 (6.6–13.4)	8.3# (5.1–11.6)	7.3* (1.0–11.6)	4.0* (3.2–5.4)
Full-length genome	9.1 (7.7–10.7)	6.6* (4.5–8.0)	6.5* (3.5–9.6)	4.3* (3.2–5.7)

[a] Fifty-six full-length HIV-1 sequences were used in the analysis, based on the hidden Markov model alignment of the entire HIV-1 genome. The following 28 sequences of subtype B were used: AUMBCC54, C18MBC, DH123, 89.6, RF, WEAU, HAN, HIVMN, BCSG3, OYI, CAM1, NY5, pNL43, LAI, HXB2, JRFL, JRCSF, AUMBC925, AUMBC200, YU2, YU10, ACH320A, ACH320B, SF2, HIV1AD8, D31, MANC, and WR27. Sequences AUMBCC54, C18MBC, and NY5 were excluded from the *nef* and 3′ LTR analysis because of deletions or the absence of sequences for these regions. Pairwise distances in four groups of sequences (pNL43, LAV, and HXB2; JRFL and JRCSF; YU2 and YU10; and ACH320A and ACH320B) were excluded from the analysis. The eight subtype B sequences from AIDS patients were JRFL, YU2, 89.6, RF, HAN, MN, BCSG3, and WR27. Thirteen sequences of subtype C were included in the analysis: 8 clones from Botswana (1 from each patient) and 5 sequences from India (301999, 21068, 301905, 301904, and 11246). All distances were calculated by DNADIST program from the PHYLIP v. 3.572 package based on hidden Markov model alignment. The transition/transversion ratios were set to 3.0 for *gag*, 1.5 for *env*, 1.42 for V3 and V1-V2, and 2.0 for all other HIV-1 genes.

[b] Statistical significance versus Botswana HIV-1 clones is shown as follows: *, $P < 0.0001$; #, $P = 0.005$; ‖, $P = 0.013$. ‡, $P = 0.028$; †, $P = 0.033$; and §, $P = 0.13$.

cause AIDS patients might be expected to have higher variability, we made the same comparison, limiting the subtype B reference to eight sequences selected from confirmed AIDS patients (column 2). The intersample diversity among Botswana clones was significantly higher than that among subtype B references or Indian samples on the level of the full-length HIV-1 genome. Across the viral genome, the mean diversity among Botswana samples was congruent with the full-length genome analysis. The intersample diversity analysis statistically confirmed the phylogenetic study observations (Fig. 1 and 2) that the newly characterized Botswana clones were highly diversified.

The results of intrasample diversity analysis were limited by the methods used (PCR amplification and cloning) and by the available number of multiple subtype B full-length clones. Seven Botswana samples (except BW12) and three subtype B sets (JRFL, YU, and ACH320) were compared across the HIV-1 genome. The range of full-length diversity among Botswana samples was 0.3% (BW17 clones) to 2.9% (BW04 clones), with an average mean value of 1.4%. Intrasample diversity showed no significant difference between subtype B and C sequences (Fig. 3). However, two concentrations of diversity (low and high) were revealed among both subtype B and C sets (Fig. 3). These concentrations of low and high diversity were distributed across the entire genome and were found to be more consistent in the structural genes (*gag*, *pol*, and *env*).

All Botswana sequences were checked for potential recombination sites by the HIV-1 Subtyping Basic BLAST (3) and by RIP (40), the results consistently showing no evidence of recombination.

Clustering with HIV-1 subtype C and the high intersample diversity were the most exceptional attributes of the 23-clone set from Botswana. A star-like shape of the subtype C cluster in the phylogenetic tree was accompanied by extremely high bootstrap values across tree branches. The topology of the phylogenetic tree suggested that a common ancestor for the Botswana sequences might have existed before the common ancestor for the Indian sequences analyzed or before the strains C-92BR025 (Brazil) and C-ETH2220 (Ethiopia) diverged.

Intersample diversity within subtype C has been previously found to vary from 5 to 11.5% (1, 7, 38). Higher levels of diversity were found among Botswana clones in this study, in spite of the fact that samples were taken from one place and at one time point. Both full-genome sequences and multiple subgenomic loci demonstrated the same patterns, with a higher mean value of variability among the Botswana samples.

The increased genetic diversity of subtype C viruses in Botswana might have different underlying causes, including a variety of host and viral factors. Among the latter factors, a combination of the genetic flexibility of subtype C virus and its multiple introductions might be the most important. A number of recent findings argue that one possible cause of the high viral diversity in the Botswana epidemic could be higher flexibility of subtype C virus and its altered ability to diversify. These arguments include, but are not limited to the following. (i) Subtype C is predominant in most recent HIV-1 epidemics worldwide (1, 7, 27, 35, 36, 38, 45, 47). (ii) The highest prevalence of HIV-1 infection in various epidemics is caused mainly by subtype C virus. (iii) Subtype C virus may have a faster disease progression (20), and patients infected with HIV-1 subtype C developed AIDS earlier than patients with subtype A virus (23). (iv) Three or four NF-κB sites (instead of two) might lead to more efficient viral transcription (13, 29, 30). (v) The TNF-α response to subtype C virus is significantly higher than to HIV-1 subtype B (28, 29), suggesting the possibility of increased viral transcription and replication in correlation with NF-κB copy number (28, 29). (vi) The viral load of subtype C infections may be higher in different compartments that might cause an increased level of viral transmission (22). On the other hand, a scenario that suggests independent diversification of the virus in other regions and delayed entry of the epidemic in Botswana, followed by multiple introductions of the subtype C virus from adjacent countries cannot be excluded (42–44).

Botswana is geographically located at the center of the AIDS epidemic in Southern Africa. UNAIDS and World Health Organization surveillance data suggest that the widespread rise of the HIV-1 epidemic in Botswana started in the early to mid-1990s and reached one of the highest prevalence rates in Africa (42–44). For more recent HIV-1 epidemics, such as those described in Thailand and India, one might expect to find a highly homogeneous pool of local viruses that formed a monophyletic phylogenetic subcluster with relatively short and aggregated branches. However the findings in this study contradict the established trend.

Extremely high interpatient diversity across the genome was supported by long branch lengths in the phylogenetic trees throughout the Botswana viruses within the genetic subtype C. No multiple subtypes or recombination were found in this study. Because it currently has the highest incidence rates of

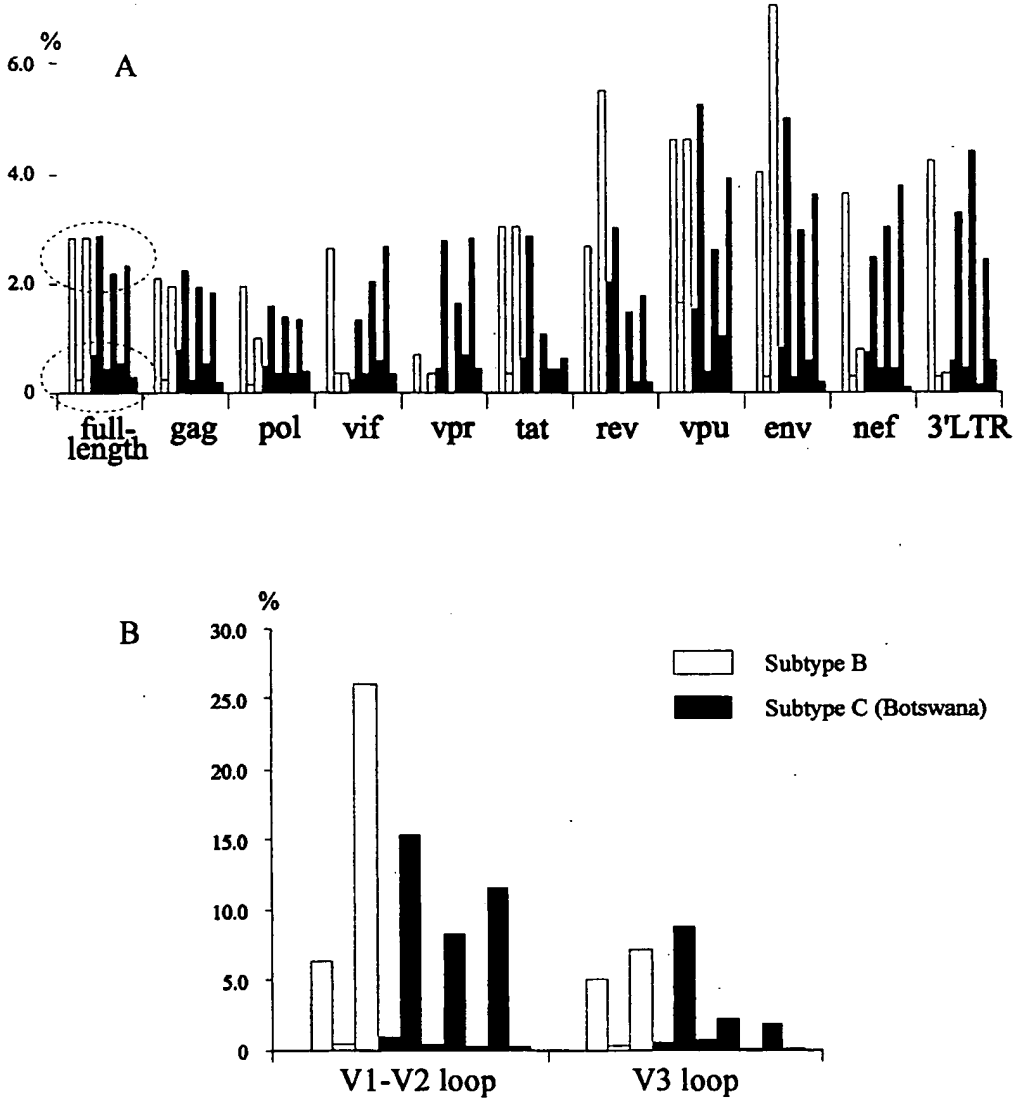


FIG. 3. Intrasample diversity among subtype B (open bars) and newly derived subtype C (solid bars) clones from Botswana. Pairwise distances were computed by the DNADIST program from the PHYLIP package based on hidden Markov model alignment of the complete genome or individual genes or subgenomic regions (B) retrieved from the same alignment. Three sets of subtype B sequences were included: JRFL and FRCSF, YU2 and YU10, and ACH320A and ACH320B. Seven sets of subtype C Botswana clones were included: BW01, BW04, BW05, BW11, BW15, BW16, and BW17. The order of bars for each gene or region corresponds to the order in which the sets are mentioned. Dashed ovals depict low- and high-diversity groups.

HIV-1, the region of southern Africa, including Botswana and Zimbabwe, is an important site for the design and development of an anti-HIV vaccine. Recent studies of cross-clade cytotoxic T-lymphocyte recognition (6, 8, 48) have made a significant contribution to AIDS vaccine development. However, the extent of cross-clade immune responses in a highly diverse HIV-1 environment and the mechanism of vaccination throughout different HLA profiles in the population are still open questions. HIV-1 strains circulating locally (or appropriate combinations of local strains) should provide a more effective prototype or candidate for a vaccine than a distinct virus circulating elsewhere. As such, a comprehensive molecular epidemiology study of locally circulating HIV-1 strains would facilitate the design of an effective AIDS vaccine for a particular population, country, or geographic region.

Nucleotide sequence accession number. The 23 full-length HIV-1 sequences from Botswana are available under GenBank accession no. AF110959 to AF110981.

This study was supported in part by OIG grant R35 CA39805-13 from the National Institutes of Health and grant D43 TW00004 from the Fogarty International Center, National Institutes of Health.

We thank Bette T. Korber for critical analysis, Suranjana Choudhury for HIV-1 culturing, Peter Gilbert for help with statistical analysis, and Rajiv Rawat for editorial assistance.

REFERENCES

1. Abebe, A., C. L. Kuiken, J. Goudsmit, M. Valk, T. Messele, T. Sahlu, H. Yeneneh, A. Fontanet, F. De Wolf, and T. F. Rinke De Wit. 1997. HIV type 1 subtype C in Addis Ababa, Ethiopia. AIDS Res. Hum. Retroviruses 13: 1071–1075.
2. Alaeus, A., T. Leitner, K. Lidman, and J. Albert. 1997. Most HIV-1 genetic

subtypes have entered Sweden. AIDS 11:199–202.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipmann.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
4. **Becker, M. L. B., G. de Jager, and W. B. Becker.** 1995. Analysis of partial *gag* and *env* gene sequences of HIV type 1 strains from Southern Africa. AIDS Res. Hum. Retroviruses 11:1265–1267.
5. **Benson, D. A., M. S. Boguski, D. J. Lipman, J. Ostell, and B. F. F. Ouellette.** 1998. GenBank. Nucleic Acids Res. 26:1–7.
6. **Betts, M. R., J. Krowka, C. Santamaria, K. Balsamo, F. Gao, G. Mulundu, C. Luo, N. N'Gandu, H. Sheppard, B. H. Hahn, S. Allen, and J. A. Frelinger.** 1997. Cross-clade human immunodeficiency virus (HIV)-specific cytotoxic T-lymphocyte responses in HIV-infected Zambians. J. Virol. 71:8908–8911.
7. **Bredell, H., C. Williamson, P. Sonnenberg, D. J. Martin, and L. Morris.** 1998. Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. AIDS Res. Hum. Retroviruses 14:677–684.
8. **Cao, H., P. Kanki, J.-L. Sankalé, A. Dieng-Sarr, G. P. Mazzara, S. A. Kalams, B. Korber, S. Mboup, and B. D. Walker.** 1997. Cytotoxic T-lymphocyte cross-reactivity among different human immunodeficiency virus type 1 clades: implications for vaccine development. J. Virol. 71:8615–8623.
9. **Centers for Disease Control and Prevention.** 1987. Centers for Disease Control and Prevention: revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. Morbid. Mortal. Weekly R.P. 36 (Suppl 1):S1–S15.
10. **Dittmar, M. T., G. Simmons, Y. Donaldson, P. Simmonds, P. R. Clapham, T. F. Schulz, and R. A. Weiss.** 1997. Biological characterization of human immunodeficiency virus type 1 clones derived from different organs of an AIDS patient by long-range PCR. J. Virol. 71:5140–5147.
11. **Eddy, S. R.** 1996. Hidden Markov models. Curr. Opin. Struct. Biol. 6:361–365.
12. **Eddy, S. R.** April 1995, posting date. HMMER. [Online.] http://hmmer.wustl.edu/. [15 October 1998, last date accessed.]
13. **Englund, G., M. D. Hoggan, T. S. Theodore, and M. A. Martin.** 1991. A novel HIV-1 isolate containing alterations affecting the NF-κB element. Virology 181:150–157.
14. **Esparza, J.** 1998. Personal communication.
15. **Essex, M.** 1998. State of the HIV pandemic. J. Hum. Virol. 1:427–429.
16. **Essex, M.** Human immunodeficiency viruses of the developing world. Semin. Virol., in press.
17. **Felsenstein, J.** 1993. PHYLIP: phylogeny inference package, 3.52c ed. University of Washington, Seattle.
18. **Fujii, S., K. Obaru, S. Matsushita, T. Morikita, H. Higuchi, K. Fujimoto, and K. Takatsuki.** 1997. Characterization of proviral DNA from an individual with long-term, nonprogressive infection with HIV-1 and nonrecoverable virus. J. Acquired Immune Defic. Syndr. Hum. Retrovirol. 15:247–256.
19. **Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barré-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1998. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J. Virol. 72:5680–5698.
20. **Hoelscher, M., D. Eimannsberger, R. Cheingsong-Popov, B. Jordan-Harder, D. Mwakagile, H. Nkwera, T. Nukulila, L. Gurtler, and F. von Sonnenburg.** 1997. A shift in age specific subtype distribution gives evidence for different pathogenicity of subtypes A and C. *In* Program of the 5th Annual HIV Dynamics & Evolution Conference, Santa Fe, N.Mex.
21. **Janssens, W., A. Buve, and J. N. Nkengasong.** 1997. The puzzle of HIV-1 subtypes in Africa. AIDS 11:705–712.
22. **John, G., J. Neilson, P. Lewis, J. Overbaugh, R. W. Nduati, D. A. Mbori-Ngacha, and J. Kreiss.** 1988. HIV-1 subtype and shedding of cervicovaginal HIV-1 DNA during pregnancy, abstr. 13117. *In* Program of the 12th World AIDS Conference, Geneva, Switzerland.
23. **Kanki, P. J., D. J. Hamel, J.-L. Sankale, C.-C. Hsieh, I. Thior, F. Barin, S. A. Woodcock, A. Gueye-Ndiaye, E. Zhang, M. Montano, T. Siby, R. Marlink, I. NDoye, M. E. Essex, and S. MBoup.** 1999. HIV-1 subtypes differ in disease progression. J. Infect. Dis. 179:68–73.
24. **Korber, B., B. Foley, T. Leitner, F. McCutchan, B. Hahn, J. W. Mellors, G. Myers, and C. Kuiken (ed.).** 1997. Human retroviruses and AIDS, p. I-1–I-127. Theoretical Biology and Biophysics, Group T-10, Los Alamos, N.Mex.
25. **Leitner, T., B. Korber, D. Robertson, F. Gao, and B. Hahn.** 1997. Updated proposal of reference sequences of HIV-1 genetic subtypes, p. III-19–III-24. *In* B. Korber, B. Foley, T. Leitner, F. McCutchan, B. Hahn, J. W. Mellors, G. Myers, and C. Kuiken (ed.), Human retroviruses and AIDS. Theoretical Biology and Biophysics, Group T-10, Los Alamos, N.Mex.
26. **Leitner, T., S. Kumar, and J. Albert.** 1997. Tempo and mode of nucleotide substitutions in *gag* and *env* gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. J. Virol. 71:4761–4770.
27. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73:152–160.
28. **Montano, M. A., C. P. Nixon, and M. Essex.** 1998. Dysregulation through the NF-κB enhancer and TATA box of the human immunodeficiency virus type 1 subtype E promoter. J. Virol. 72:8446–8452.
29. **Montano, M. A., C. P. Nixon, T. P. Ndung'u, A. V. A. Novitsky, H. Bussman, D. Dickman, S. M. Gaolekwe, and M. Essex.** Transcriptional gain-of-function in TNF-α mediated activation of HIV-1 subtype C. Submitted for publication.
30. **Montano, M. A., V. A. Novitsky, J. T. Blackard, N. L. Cho, D. A. Katzenstein, and M. Essex.** 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J. Virol. 71:8657–8665.
31. **Orloff, G. M., M. Kalish, L. J. Chiphangwi, K. E. Potts, C.-Y. Ou, G. Schochetman, G. Dallabetta, A. I. Saah, and P. G. Miotti.** 1993. V3 loops of HIV-1 specimens from pregnant women in Malawi uniformly lack a potential N-linked glycosylation site. AIDS Res. Hum. Retroviruses 9:705–706.
32. **Page, R. D. M.** 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comp. Appl. Biosci. 12:357–358.
33. **Perriere, G., and M. Gouy.** 1996. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78:364–369.
34. **Proutski, V., and E. Holmes.** 1998. SWAN: sliding window analysis of nucleotide sequence variability. Bioinformatics 14:467–468.
35. **Ray, S., K. S. Lole, R. S. Paranjape, S. Kulkarni, D. Gadkari, H. Sheppard, and R. C. Bollinger.** 1998. Cloning of full-length HIV-1 genomes from recent infections in India: recombination and CTL epitope, abstr. 11194. *In* Program of the 12th World AIDS Conference, Geneva, Switzerland.
36. **Renjifo, B., B. Chaplin, D. Mwakagile, P. Shah, F. Vannberg, G. Msamanga, D. Hunter, W. Fawzi, and M. Essex.** 1998. Epidemic expansion of HIV type 1 subtype C and recombinant genotypes in Tanzania. AIDS Res. Hum. Retroviruses 14:635–638.
37. **Salminen, M. O., J. K. Carr, D. L. Robertson, P. Hegerich, D. Gotte, C. Koch, E. Sanders-Buell, F. Gao, P. M. Sharp, B. H. Hahn, D. S. Burke, and F. E. McCutchan.** 1997. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. J. Virol. 71:2647–2655.
38. **Salminen, M. O., B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. Leinikki, D. S. Burke, and F. E. McCutchan.** 1996. Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. AIDS Res. Hum. Retroviruses 12:1329–1339.
39. **Salminen, M. O., C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. Virology 213:80–86.
40. **Siepel, A. C., A. L. Halpern, C. Macken, and B. T. Korber.** 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. AIDS Res. Hum. Retroviruses 11:1413–1416.
41. **Trinh, T., J. Jesse, F. R. Bloom, and V. Hirsch.** 1994. STBL2TM: an Escherichia coli strain for the stable propagation of retroviral clones and direct repeat sequences. Focus 16:78–80.
42. **UNAIDS, and WHO.** 22 June 1998, posting date. Global HIV/AIDS & STD Surveillance. Epidemiological fact sheets by country. [Online.] http://www.us.unaids.org/highband/document/epidemio/june98/fact_sheets/index.html. [15 October 1998, last date accessed.]
43. **UNAIDS, and WHO.** 1998. Report on the Global HIV/AIDS Epidemic—June 1998. UNAIDS and WHO, Geneva, Switzerland.
44. **UNAIDS, and WHO.** 29 June 1998, posting date. The status and trends of the HIV/AIDS epidemic in the world. [Online.] http://www.us.unaids.org/highband/document/epidemio/map.pdf,Geneva. [15 October 1998, last date accessed.]
45. **van Harmelen, J., H. Bredell, L. Morris, E. van der Ryst, S. Lyons, D. York, and C. Williamson.** 1998. Determination of prevalent HIV-1 subtypes in South Africa and the construction of a gp120 DNA vaccine, abstr. 33218. *In* Program of the 12th World AIDS Conference, Geneva, Switzerland.
46. **van Harmelen, J., R. Wood, M. Lambrick, E. P. Rybicki, A. L. Williamson, and C. Williamson.** 1997. An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. AIDS 11:81–87.
47. **Williamson, C., S. Engelbrecht, M. Lambrick, E. van Rensburg, R. Wood, W. Bredell, and A.-L. Williamson.** 1995. HIV-1 subtypes in different risk groups in South Africa. Lancet 346:782.
48. **Wilson, S. E., S. L. Pedersen, J. C. Kunich, V. L. Wilkins, D. L. Mann, G. P. Mazzara, J. Tartaglia, C. L. Celum, and H. W. Sheppard.** 1998. Cross-clade envelope glycoprotein 160-specific CD8+ cytotoxic T lymphocyte responses in early HIV type 1 clade B infection. AIDS Res. Hum. Retroviruses 14:925–937.

JOURNAL OF VIROLOGY, Jan. 2002, p. 397–405
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.1.397–405.2002

Vol. 76, No. 1

Molecular Epidemiology of Human Immunodeficiency Virus Type 1 Transmission in a Heterosexual Cohort of Discordant Couples in Zambia

Stanley A. Trask,[1] Cynthia A. Derdeyn,[2] Ulgen Fideli,[3] Yalu Chen,[1] Sreelatha Meleth,[1] Francis Kasolo,[4] Rosemary Musonda,[5] Eric Hunter,[2] Feng Gao,[1] Susan Allen,[3] and Beatrice H. Hahn[1,2]*

Departments of Medicine[1] and Microbiology,[2] School of Medicine, and Department of Epidemiology and International Health, School of Public Health,[3] University of Alabama at Birmingham, Birmingham, Alabama 35294, and Department of Pathology and Microbiology, University Teaching Hospital, Lusaka,[4] and Tropical Disease Research Center, Ndola,[5] Zambia

Received 24 July 2001/Accepted 24 September 2001

Most human immunodeficiency virus type 1 (HIV-1) transmissions in sub-Saharan Africa are believed to occur between married adults who are discordant for their HIV-1 infection status; however, no studies to date have investigated the molecular epidemiology of such transmission events. Here we report the genetic characterization of HIV-1 strains from 149 transmission pairs that were identified prospectively in a cohort of discordant couples in Lusaka, Zambia. Subgenomic *gag*, gp120, gp41, and/or long terminal repeat regions were amplified by PCR analysis of uncultured blood samples from both partners and sequenced without interim cloning. Pairwise genetic distances were calculated for the regions analyzed and compared to those of subtype-specific reference sequences as well as local controls. Sequence relationships were also examined by phylogenetic tree analysis. By these approaches, epidemiological linkage was established for the majority of transmission pairs. Viruses from 129 of the 149 couples (87%) were very closely related and clustered together in phylogenetic trees in a statistically highly significant manner. In contrast, viruses from 20 of the 149 couples (13%) were only distantly related in two independent genomic regions, thus ruling out transmission between the two partners. The great majority (95%) of transmitted viruses were of subtype C origin, although representatives of subtypes A, D, G, and J were also identified. There was no evidence for extensive transmission networks within the cohort, although two phylogenetic subclusters of viruses infecting two couples each were identified. Taken together, these data indicate that molecular epidemiological analyses of presumed transmission pairs are both feasible and required to determine behavioral, virological, and immunological correlates of heterosexual transmission in sub-Saharan Africa with a high level of accuracy.

By the end of the year 2000, an estimated 36 million adults and children were living with human immunodeficiency virus (HIV) infection-AIDS worldwide (39). More than 70% of these individuals resided in sub-Saharan Africa, where the average prevalence of HIV infection is currently 8.8% and transmissions occur predominantly through heterosexual routes or from mother to child (30). One of the African countries with a particularly high prevalence of human immunodeficiency virus type 1 (HIV-1) infection is Zambia, where it is estimated that 20% of all adults harbor HIV-1 and 20% of all cohabitating couples are discordant for their HIV-1 infection status (i.e., one partner is HIV-1 positive and the other is negative) (40). Novel interventions designed to curtail the explosive spread of HIV-1 in Zambia and other high-prevalence countries in sub-Saharan Africa are thus urgently needed but are likely to require detailed knowledge about the factors that influence heterosexual transmission.

The Zambia-UAB HIV Research Project (ZUHRP) was established in 1994 to provide voluntary HIV-1 testing and counseling, long-term monitoring, and health care to cohabitating couples in the capital city of Lusaka (3, 25). To date, 9,569 couples have been tested for HIV-1, of whom 21% were HIV-1 discordant, 26% were concordant HIV-1 positive, and 53% were concordant HIV-1 negative at the time of enrollment. Between February 1994 and October 2000, 1,022 discordant couples (535 with HIV-1-infected men and 487 with HIV-1-infected women) were enrolled into a prospective study of the incidence and predictors of heterosexual transmission and were monitored at 3-month intervals for seroconversion of the seronegative partner. Although testing and counseling prompted substantial risk reduction in this cohort, a seroconversion rate of 8.5 per 100 person years remained, which was similar for male-to-female and female-to-male transmissions (12). Because frequent follow-up visits facilitated blood collection from both the putative donor and the recipient after a transmission event, this cohort has provided a unique setting to examine the incidence, demographics, and behavioral and biological correlates as well as the viral and host determinants of heterosexual transmission of HIV-1. However, a prerequisite for the acquisition of meaningful data, particularly with regard to predictors of contagion in the index seropositive partner, is

* Corresponding author. Mailing address: Department of Medicine, University of Alabama at Birmingham, 720 20th Street South, KAUL 816, Birmingham, AL 35294. Phone: (205) 934-0412. Fax: (205) 934-1580. E-mail: bhahn@uab.edu.

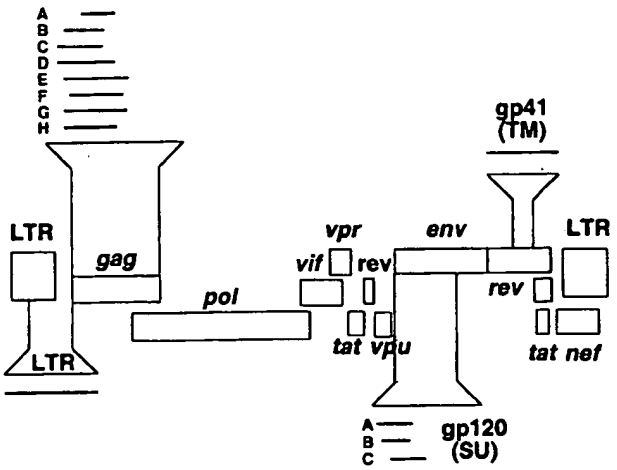


FIG. 1. HIV-1 subgenomic regions utilized for epidemiological linkage analysis. A schematic representation of the HIV-1 genome is shown, with brackets denoting the subgenomic *gag*, gp120, gp41, and LTR fragments that were amplified for diagnostic PCR analysis. Overlapping *gag* and gp120 regions are denoted by capital letters (*gagA* to *-H*; gp120A to -C) and are referred to in Tables 2 to 4.

the ability to confirm, with a high level of confidence, epidemiological linkage of HIV-1 transmission between members of all putative transmission pairs.

Molecular analyses of suspected transmission links have been widely used to characterize localized HIV-1 outbreaks, mother-to-infant transmission, sexual transmission, sharing of contaminated needles, donation of contaminated blood, receipt of contaminated clotting reagent, nosocomial transmissions from health care workers, and intrafamilial contacts (1, 4, 5, 6, 13, 18, 21, 29, 36, 42, 44). In all of these cases, the establishment of epidemiological linkage relied on the documentation of closer genetic relatedness between viruses infecting the suspected transmission pair(s) compared to control viruses isolated from unrelated individuals in the same region. Here we developed a similar approach to confirm (or refute) heterosexual transmission among discordant couples within the ZUHRP cohort.

Blood samples were collected between 1996 and 2000 from both partners of 149 (of a total of 162) discordant couples in whom seroconversion had been documented. The time period between the last negative and the first positive blood tests for the seroconvertor (which in most cases was also the blood sample used for linkage analysis) was 4.9 months on average, but in some cases it extended up to 4 years. High-molecular-weight DNA was extracted from whole blood or Ficoll gradient-purified peripheral blood mononuclear cells by using the QIAamp Blood Kit (Qiagen, Valencia, Calif.). For a small number of couples, DNA was extracted from dried blood spots (9). Because of the known variability of HIV-1, different regions of the HIV-1 genome were targeted for PCR amplification, resulting in comparisons of *gag*, gp120, gp41, and/or long terminal repeat (LTR) regions as shown in Fig. 1. Although the gp41 primers were by far the most cross-reactive, the suitability of this primer set was discovered only after alternative genomic regions from a number of transmission pairs had already been analyzed (43). LTR, gp120, and gp41 primers and amplification conditions have been described previously (15, 16). The primers that were used to amplify sequences within *gag* were cgagA 5′-TGATAAAACCTCCAATTCCCCCTA T-3′ and PBS1A 5′-TTTGCCTGTACTGGGTCTCTCTGGT T-3′ in the first round and cgagB 5′-AATACTGTATCATCT GCTCCTGTATC-3′ and PBS1B 5′-GCTTAAGCCTCAATA AAGCTTGCCTT-3′ in the second round. PCR products were sequenced directly, using cycle sequencing and dye terminator methodologies, on an automated DNA sequencer (model 377A; Applied Biosystems, Inc., Foster City, Calif.). Both strands of the PCR products were sequenced (sequences are available under GenBank accession numbers AF404868 through AF405203, AF406742, and AF406743). Although population-based sequencing was used to allow analysis of the predominant viral form in each individual, the number of ambiguous base pairs in the entire data set was <0.3%.

To establish suitable linkage criteria for HIV-1 strains infecting the Zambian couples, amplified viral sequences were first subjected to preliminary phylogenetic tree analyses to identify all circulating HIV-1 group M subtypes (not shown). Full-length and nonrecombinant reference sequences representing these subtypes were then obtained from the Los Alamos HIV Sequence Database (Table 1) and subjected to pairwise sequence comparisons in the genomic regions corre-

TABLE 1. Subtype-specific full-length reference sequences from the HIV Sequence Database

Subtype_sequence name	Accession no.	Reference
A_Q2317	AF004885	31
A_SE8891	AF069673	7
A_SE8538	AF069669	7
A_SE6594	AF069672	7
A_SE7535	AF069671	7
A_SE7253	AF069670	7
A_SE8131	AF107771	7
A_U455	M62320	28
A_92UG037.1	U51190	14
C_96BW04.07	AF110963	26
C_96BW11B01	AF110971	26
C_96BW15C02	AF110974	26
C_96BW05.04	AF110968	26
C_96BW16.26	AF110978	26
C_96BW12.10	AF110972	26
C_96BW17A09	AF110979	26
C_96BW01B21	AF110960	26
C_C2220	U46016	35
C_94IN11246	AF067159	24
C_98BR004	AF286228	33
C_98IS002.5	AF286233	33
C_98TZ013.10	AF286234	33
C_98TZ017.2	AF286235	33
C_97ZA012.1	AF286227	33
D_94UG114.1	U88824	14
D_84ZR085	U88822	14
D_NDK	M27323	37
D_ELI	K03454	2
D_Z2Z6	M22639	38
G_DRCBL	AF084936	27
G_HH8793	AF061641	8
G_92NG083.2	U88826	14
G_SE6165	AF061642	8
J_SE91733	AF082395	23
J_SE92809	AF082394	23

TABLE 2. Genetic diversity in different subgenomic regions for two sets of reference sequences

Subgenomic region[a]	Sequence subtype	Reference sequences[b]				Zambian donor sequences[b]			
		n	Mean distance	SD	Cutoff value	*n*	Mean distance	SD	Cutoff value
gagA	C	15	5.5	1.5	2.5	6	5.5	1.4	2.7
gagB	C	15	6.9	1.3	4.3	12	5.6	1.3	3.0
gagC	C	15	7.7	1.6	4.5	7	6.3	1.3	3.7
gagD	A	9	6.7	1.8	3.1	3	13.1	1.6	9.9
gagD	C	15	7.6	1.5	4.6	6	8.4	2.5	3.4
gagE	C	15	6.1	1.3	3.5	7	5.3	2.1	1.1
gagF	C	15	6.0	1.4	3.2	8	4.2	0.8	2.6
gagG	C	15	6.2	1.3	3.6	9	5.4	0.6	4.2
gagH	C	15	6.3	1.3	3.7	4	7.9	1.0	5.9
gp120A	C	15	8.6	1.8	5.0	4	6.5	1.1	4.3
gp120B	C	15	7.0	1.9	3.2	11	5.9	1.1	3.7
gp120C	C	15	13.1	2.4	8.3	3	13.0	1.9	9.2
gp41	C	15	8.4	1.5	5.4	81	9.3	1.9	5.5
gp41	D	5	5.4	1.4	2.6	1	NA	NA	NA
gp41	G	4	6.8	1.5	3.8	3	5.6	2.0	1.6
gp41	J	2	NA	NA	NA	1	NA	NA	NA
LTR	C	NA	NA	NA	NA	3	3.2	0.4	2.4

[a] The subgenomic region was analyzed as shown in Fig. 1.
[b] Full-length (nonmosaic) subtype-specific reference sequences were obtained from the Los Alamos Sequence Database and are listed in Table 1. Mean distance is the mean percent sequence difference in the analyzed genomic region. *n*, number of sequences compared. The cutoff value was 2 SD below the mean. NA, not available.

sponding to the PCR amplification products. Eight partially overlapping regions in *gag*, three in gp120, one in gp41, and one in the LTR were used for analysis (Fig. 1). Uncorrected nucleotide sequence distances were then calculated for each transmission pair and compared to the mean sequence distances calculated for the reference sequence set in the corresponding genomic region. The latter minus two standard deviations (SDs) was arbitrarily assigned as the cutoff value for epidemiologically linked sequence pairs (Table 2). Transmission pairs were tentatively classified as epidemiologically linked when their pairwise sequence distances fell below this limit. Conversely, transmission pairs were tentatively classified as unlinked when their pairwise distances exceeded this limit. For the subtype C reference set, only single representatives from India and Brazil were included, so as to not skew results due to the more recent introduction of HIV-1 into these countries. Mean distances for the gp41 region of subtype J and the LTR region of subtype C could not be calculated because of a lack of sufficient reference sequences.

Although the HIV-1 epidemic in Zambia is believed to be longstanding and mature (33), we examined the extent of genetic diversity of HIV-1 strains infecting all putative Zambian donors to exclude the possibility of a recent founder effect within this cohort. As shown in Table 2, pairwise comparison of all Zambian donor sequences yielded mean distance values, SDs, and cutoff values that were very similar to those obtained for the reference sequences. This indicated that the selected reference sequences were indeed representative of the viruses infecting the cohort. There was no evidence for an unusually high degree of genetic relatedness among the Zambian donor viruses that could have confounded the linkage analysis. Instead, the results suggested that the viruses circulating within the heterosexual transmission cohort were representative of the viruses circulating in the country at large.

Having established suitable reference sequence sets, we next used the linkage criteria (Table 2) to tentatively classify the 149 transmission pairs as either likely linked or unlinked. Table 3 lists the identification number, dates of blood collection from donor and recipient (identical unless indicated otherwise), genomic region analyzed, and viral subtype for 129 transmission pairs whose uncorrected pairwise distances fell below the cutoff value of the reference sequences (compare with Table 2). Only one transmission pair (couple 136) yielded a pairwise distance (2.7%) that was slightly above the reference cutoff limit (2.6%). However, this pair was included as a likely linked transmission event after inspection of the two sequences revealed G-to-A hypermutation (41) as the cause of 9 of 10 sequence changes between donor and recipient virus. G-to-A hypermutation was also identified as a reason for increased genetic diversity in four other pairs (couples 65, 132, 138, and 149), although in these instances distance values did not exceed the cutoff limit. The majority of all couples listed in Table 3 (127 of 129) also fell below the cutoff value of the Zambian donor sequences. These data thus indicated that most couples harbored viruses whose sequences were considerably more homologous to one another than to unrelated reference sequences from the database as well as local controls.

Distance calculations also identified 20 couples harboring HIV-1 strains whose uncorrected pairwise distances exceeded the corresponding cutoff values, and this was confirmed by sequencing two independent genomic regions (Table 4). The great majority of pairwise distances from these transmission pairs fell well above the cutoff values of both sets of reference sequences (compare with Table 2), thus indicating a clearly discernible difference between linked and unlinked transmission pairs (in the LTR region, Zambian donor sequences served as the sole reference set). This is best illustrated in Fig. 2, where the pairwise distances of 15 subtype C reference sequences in the gp41 region are contrasted to the corresponding gp41 distances from 66 linked and 15 unlinked (subtype C) cohort transmission pairs. The median sequence distance of the viral group tentatively classified as linked was significantly different from the median distance of the viral group tentatively classified as unlinked as well as the median distance of

TABLE 3. Genetic distances for linked Zambian transmission pairs

Sample ID	Sample collection date (mo-day-yr)	Subgenomic region[a]	Cutoff value[b]	Pairwise distance[c]	Subtype
47	6-5-98	*gagA* (242)	2.5	0.0	C
67	6-8-98	*gagA*	2.5	0.4	C
82	10-19-98	*gagA*	2.5	1.7	C
89	6-13-98	*gagA*	2.5	0.4	C
114	8-2-98	*gagA*	2.5	0.8	C
119	8-9-98	*gagA*	2.5	0.0	C
4	11-15-98	*gagB* (285)	4.3	1.1	C
11	6-6-98	*gagB*	4.3	1.1	C
23	8-7-98	*gagB*	4.3	0.4	C
16	2-10-98	*gagB*	4.3	0.7	C
17	8-9-96	*gagB*	4.3	0.0	C
25	6-12-97	*gagB*	4.3	2.5	C
29	6-7-98	*gagB*	4.3	4.0	C
33	6-8-98	*gagB*	4.3	0.7	C
42	7-22-96 (F) 8-9-96 (M)	*gagB*	4.3	0.0	C
49	7-9-96	*gagB*	4.3	0.4	C
1	10-16-98	*gagC* (317)	4.5	2.3	C
2	7-4-98	*gagC*	4.5	2.0	C
5	6-15-98	*gagC*	4.5	0.6	C
10	2-14-96	*gagC*	4.5	1.3	C
30	6-8-98	*gagC*	4.5	0.0	C
71	9-5-98	*gagC*	4.5	1.7	C
20	10-15-98	*gagD* (406)	3.1	1.5	A
56	6-6-98	*gagD*	3.1	1.5	A
142	8-13-98	*gagD*	3.1	1.0	A
46	6-7-98	*gagD*	4.6	0.5	C
55	8-13-98	*gagD*	4.6	0.8	C
62	6-6-98	*gagD*	4.6	2.2	C
103	7-8-98	*gagD*	4.6	1.1	C
76	6-25-98	*gagE* (458)	3.5	0.9	C
101	6-20-98	*gagE*	3.5	0.4	C
108	6-19-98 (F) 6-20-98 (M)	*gagE*	3.5	1.1	C
124	8-15-98	*gagE*	3.5	0.2	C
37	6-8-98	*gagF* (381)	3.2	2.1	C
51	6-6-98	*gagF*	3.2	2.4	C
58	6-19-98	*gagF*	3.2	0.3	C
64	8-6-98	*gagF*	3.2	1.1	C
70	8-27-98	*gagF*	3.2	1.1	C
68	8-9-98	*gagG* (445)	3.6	2.5	C
92	6-8-98	*gagG*	3.6	0.9	C
97	6-6-98	*gagG*	3.6	2.9	C
99	6-6-98	*gagG*	3.6	0.2	C
102	6-23-98	*gagG*	3.6	1.4	C
105	7-21-98	*gagG*	3.6	1.4	C
106	6-8-98	*gagG*	3.6	1.1	C
110	8-20-98	*gagG*	3.6	0.5	C
22	6-7-98	*gagH* (373)	3.7	3.1	C
73	6-6-98	*gagH*	3.7	0.8	C
21	8-21-98 (F) 5-31-98 (M)	gp120A (373)	5.0	2.5	C
60	6-6-98	gp120A	5.0	0.6	C
113	8-7-98	gp120A	5.0	1.9	C
131	8-23-98	gp120A	5.0	0.0	C
66	11-6-98	gp120B (296)	3.2	0.3	C
75	8-7-98	gp120B	3.2	1.7	C
78	6-30-98	gp120B	3.2	0.3	C
83	8-10-98	gp120B	3.2	1.0	C
90	6-6-98	gp120B	3.2	0.7	C
122	11-12-98	gp120B	3.2	1.0	C
136[d]	10-13-99	gp41 (411)	2.6	2.7[e]	D
43	3-16-97 (F) 6-24-98 (M)	gp41	3.8	1.0	G
91	6-6-98	gp41	3.8	2.3	G
148	3-4-00	gp41	3.8	0.5	G
98[d]	8-19-97 (F) 10-26-98 (M)	gp41	n/a	3.8	J
3	9-9-99	gp41	5.4	0.5	C
6	2-16-00	gp41	5.4	1.5	C
8	7-29-96	gp41	5.4	2.5	C
12	6-6-00	gp41(411)	5.4	1.8	C
13	9-12-99	gp41	5.4	0.8	C
14	10-28-99	gp41	5.4	0.6	C
15	9-23-99	gp41	5.4	1.8	C
19	6-6-98	gp41	5.4	2.3	C
24	7-13-96	gp41	5.4	1.8	C
26	11-4-99	gp41	5.4	0.0	C
27	4-15-97	gp41	5.4	1.1	C
28	5-16-96	gp41	5.4	0.5	C
32	8-6-99	gp41	5.4	0.8	C
34	6-14-96 (F) 5-14-96 (M)	gp41	5.4	1.5	C
35	12-12-97 (F) 10-12-96 (M)	gp41	5.4	2.0	C
36	8-25-99	gp41	5.4	0.3	C
38	4-8-00	gp41	5.4	1.8	C
39	8-27-99	gp41	5.4	1.1	C
40	7-25-97 (F) 10-16-98 (M)	gp41	5.4	2.5	C
41	10-20-99	gp41	5.4	2.8	C
45	3-29-00	gp41	5.4	2.2	C
48	9-3-99 (F) 9-19-00 (M)	gp41	5.4	1.5	C
50	9-5-96	gp41	5.4	1.5	C
53	2-12-00	gp41	5.4	0.8	C
59	9-2-96	gp41	5.4	0.6	C
61	9-28-98	gp41	5.4	2.4	C
63	4-13-00	gp41	5.4	1.8	C
65	11-7-99	gp41	5.4	4.8	C
74	8-8-98	gp41	5.4	1.7	C
77	4-30-97	gp41	5.4	0.5	C
79	9-3-99 (F) 6-30-98 (M)	gp41	5.4	1.3	C
80	1-14-00	gp41	5.4	1.3	C
81	2-11-98	gp41	5.4	1.5	C
84	11-7-99	gp41	5.4	1.8	C
85	10-8-97 (F) 5-29-97 (M)	gp41	5.4	2.7	C
86	11-12-98	gp41	5.4	0.8	C
93	10-18-99	gp41	5.4	0.5	C
95	9-12-00	gp41	5.4	2.0	C
100	6-30-00	gp41	5.4	0.5	C
107	8-11-00 (F) 8-16-00 (M)	gp41	5.4	1.5	C
109	3-16-00	gp41	5.4	2.2	C
111	9-15-00	gp41	5.4	2.0	C
112	7-5-00	gp41	5.4	0.5	C
115	1-13-00	gp41	5.4	0.8	C
116	6-7-98	gp41	5.4	0.3	C
117	10-2-97	gp41	5.4	2.6	C
118	10-31-99	gp41	5.4	1.8	C
120	11-5-99	gp41	5.4	1.0	C
125	4-14-00	gp41	5.4	0.8	C
127	10-31-99	gp41	5.4	1.8	C
128	9-25-98	gp41	5.4	2.2	C
129	8-13-99	gp41	5.4	2.3	C
130	6-16-00 (F) 9-14-00 (M)	gp41	5.4	1.0	C
132	8-11-00	gp41	5.4	3.5	C
134	1-12-00	gp41	5.4	3.8	C
135	6-28-98	gp41	5.4	2.5	C
137	10-10-99	gp41	5.4	2.5	C
138	10-2-98	gp41	5.4	3.3	C
139	8-2-00	gp41	5.4	1.3	C
140	11-11-99	gp41	5.4	1.3	C
141	8-13-99	gp41	5.4	1.5	C
143	8-26-99	gp41	5.4	0.8	C
145	8-18-99	gp41	5.4	0.8	C
146	8-23-00	gp41	5.4	0.8	C
147	10-23-99	gp41	5.4	0.5	C
149	9-30-00	gp41	5.4	3.8	C

[a] The subgenomic region was analyzed as shown in Fig. 1. The subgenomic region size (in base pairs) is given in parentheses. F, female partner; M, male partner.
[b] The cutoff value is 2 SD below the mean (see Table 2).
[c] Pairwise distances are the percent sequence differences in the analyzed genomic region.
[d] Determined as epidemiologically linked by phylogenetic tree analysis.
[e] Genetic distance is primarily due to G-to-A hypermutation.

TABLE 4. Genetic distances for unlinked Zambian transmission pairs

Sample ID	Sample collection date (mo-day-yr)	Subgenomic region[a]	Cutoff value[b]	Pairwise distance[c]	Subtype
7	9-23-98	*gagB*	4.3	5.7	C
	9-23-98	gp120C	8.3	11.1	C
9	8-16-00	gp41	5.4	11.2	C
	8-16-00	*gagF*	3.2	5.9	C
18	10-19-98	*gagB*	4.3	7.5	C
	10-19-98	gp120C	8.3	12.1	C
31	2-29-00	gp41	5.4	8.5	C
	2-29-00	*gagE*	3.5	7.5	C
44	11-11-98 (F)	gp120C	8.3	16.9	C
	8-11-98 (M)	gp41	5.4	7.7	C
52	6-7-98	*gagH*	3.7	6.8	C
	6-7-98	gp120B	3.2	6.6	C
54	11-17-99	gp41	5.4	7.8	C
	11-17-99	*gagF*	3.2	4.0	C
57	5-29-97	gp41	5.4	11.0	C
	5-29-97	LTR	2.4[d]	3.8	C
69	6-7-98	*gagD*	4.6	11.0	C
	6-7-98	gp120B	3.2	3.7	C
72	6-12-98	gp120B	3.2	5.8	C
	6-12-98	gp41	5.4	9.0	C
87	6-6-00	gp41	5.4	7.5	C
	6-6-00	*gagE*	3.5	6.8	C
88	8-22-99 (F)	gp41	5.4	8.5	C
	9-19-00 (M)	*gagC*	4.5	9.8	C
94	10-25-98	*gagD*	4.6	7.6	C
	10-25-98	gp41	5.4	12.4	C
96	6-12-98 (F)	gp41	5.4	13.1	C
	2-3-99 (M)	LTR	2.4[d]	3.1	C
104	6-26-98	gp41	5.4	9.5	C
	6-26-98	LTR	2.4[d]	3.8	C
121	4-10-98	*gagH*	3.7	10.4	C
	4-10-98	gp120B	3.2	5.2	C
123	6-26-00	*gagG*	3.6	7.0	C
	6-26-00	gp41	5.4	6.8	C
126	10-27-98	gp41	5.4	8.4	C
	10-27-98	gp120B	3.2	4.1	C
133	3-16-00	gp41	5.4	12.2	C
	3-16-00	*gagE*	3.5	5.6	C
144	7-29-00	gp41	5.4	8.8	C
	7-29-00	*gagF*	3.2	3.7	C

[a] Values are as defined in Table 3, footnote *a*. The subgenomic region was analyzed as shown in Fig. 1. F, female partner; M, male partner.

[b] The cutoff value was 2 SD below the mean of the reference sequence set (see Table 2).

[c] Pairwise distances are the percent sequence differences in the analyzed genomic region.

[d] The cutoff value for LTR sequences was derived from Zambian donor sequences.

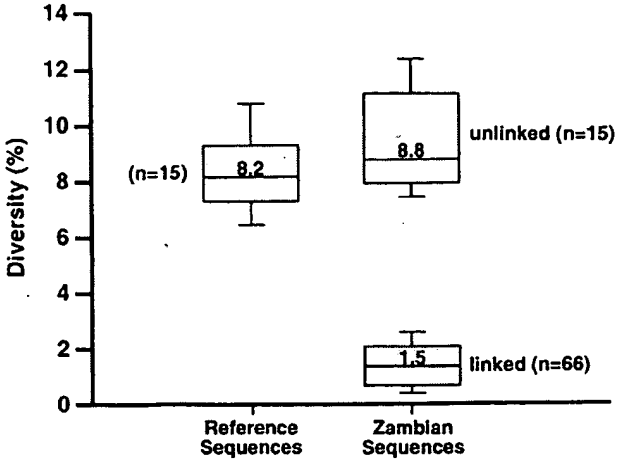
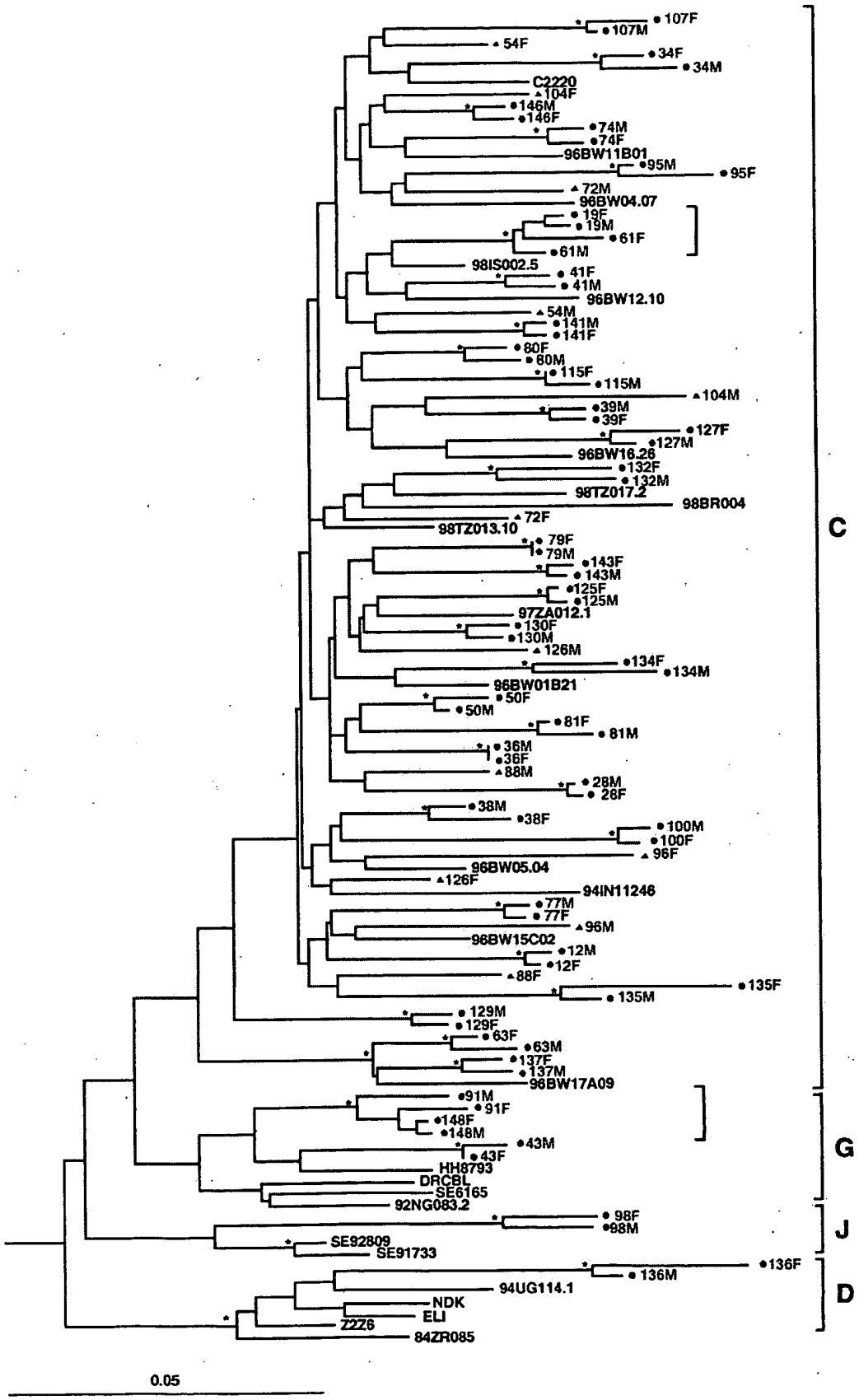


FIG. 2. Within-group diversity among linked and unlinked Zambian transmission pairs and corresponding reference sequences. Subtype C reference sequences (n = 15) from the Los Alamos HIV/SIV Sequence Database (Table 1) were subjected to pairwise sequence comparisons in the region corresponding to the PCR-amplified gp41 fragment shown in Fig. 1. Pairwise sequence distances were also calculated for 66 subtype C transmission pairs classified as linked and 15 subtype C transmission pairs classified as unlinked in the same genomic region. The distribution of distance values for these three different groups is depicted as boxes, with the lower and upper limits of the box delineating the 25th and 75th percentiles and the bars indicating the 10th and 90th percentiles, respectively. The median distance of the linked viral group (median = 1.5) was significantly different from that of both the unlinked viral group (median distance = 8.8) and the reference sequence group (median distance = 8.2) ($P < 0.0001$, one-sided Mann-Whitney test [17]). In contrast, the median sequence distance of the unlinked viral group was not statistically different from that of the reference sequence group ($P > 0.05$, Mann-Whitney test).

the reference sequence group ($P < 0.0001$) by using a one-sided Mann-Whitney test (17). In contrast, the median sequence distance of the unlinked viral group was not statistically different from that of the reference sequences ($P > 0.05$, Mann-Whitney test).

In a final set of experiments, epidemiological linkage was assessed by phylogenetic tree analysis. PCR-derived viral sequences from both partners were added to an existing master alignment (obtained from the Los Alamos HIV/SIV Sequence Database [http://hiv-web.lanl.gov/HTML/alignments.html]) that contained all reference sequences listed in Table 1. Sequences were aligned by using CLUSTAL W (19) and adjusted manually by using MASE (10). Sites with a gap in any of the sequences or sites that were ambiguous due to the population sequence approach were excluded from further analyses. Evolutionary distances were corrected for superimposed hits by using Kimura's two-parameter method (22). Phylogenetic trees were constructed by using the neighbor-joining method (34), and the reliability of topologies was estimated by using the bootstrap approach (11). Bootstrap values of ≥80% were considered significant (4, 20, 29, 36, 44). An example of a phylogenetic tree constructed from gp41 sequences of 42 transmission pairs and 26 reference sequences is shown in Fig. 3. All transmission pairs initially classified as linked by pairwise distance analysis (depicted in red) also clustered together in phylogenetic trees with significant bootstrap values (indicated by asterisks). Similarly, all transmission pairs initially classified as unlinked (depicted in blue) were not significantly related in phylogenetic trees. The latter was true for the two independent genomic regions analyzed (not shown). Finally, viral sequences derived from 98M and 98F (Fig. 3), which clustered with subtype J viruses, were significantly related to each other and thus classified as epidemiologically linked.

Phylogenetic tree analysis also yielded a subtype designation for each of the viruses infecting the 149 transmission pairs (data not shown). As shown in Fig. 3, the overwhelming majority (141 of 149; 95%) of enrolled couples were infected with subtype C viruses. Three couples harbored subtype G viruses,


107F
107M
54F
34F
34M
C2220
104F
146M
146F
74M
74F
96BW11B01
95M
95F
72M
96BW04.07
19F
19M
61F
61M
98IS002.5
41F
41M
96BW12.10
54M
141M
141F
80F
80M
115F
115M
104M
39M
39F
127F
127M
96BW16.26
132F
132M
98TZ017.2
98BR004
72F
98TZ013.10
79F
79M
143F
143M
125F
125M
97ZA012.1
130F
130M
126M
134F
134M
96BW01B21
50F
50M
81F
81M
36M
36F
88M
28M
28F
38M
38F
100M
100F
96F
96BW05.04
126F
94IN11246
77M
77F
96M
96BW15C02
12M
12F
88F
135F
135M
129M
129F
63F
63M
137F
137M
96BW17A09
91M
91F
148F
148M
43M
43F
HH8793
DRCBL
SE6165
92NG083.2
98F
98M
SE92809
SE91733
136F
136M
94UG114.1
NDK
ELI
Z2Z6
84ZR085
C
G
J
D
0.05

three couples harbored subtype A viruses, one couple harbored subtype D viruses, and one couple harbored subtype J viruses, all representing linked transmissions. To determine whether non-subtype C viruses were introduced more recently, patient records were examined for the first occurrence of non-subtype C viruses (not shown). The results revealed no particular association between the date of enrollment and the appearance of non-subtype C strains within the ZUHRP cohort: couples infected with subtype A viruses were enrolled in 1996 and 1999; couples infected with subtype G were enrolled in 1995, 1996, and 1998; couples infected with subtype D were enrolled in 1998; and couples infected with subtype J were enrolled in 1997. If we assume no recombination in the remainder of the genome, these results indicate that subtype C predominates within the ZUHRP cohort.

Finally, phylogenetic analysis allowed us to examine the evolutionary history of the cohort viruses compared to other HIV-1 strains from the same subtype. In particular, we were interested in determining whether ZUHRP couples were participating in transmission networks involving closely related viruses. Inspection of the phylogenetic tree in Fig. 3 revealed only two significant subclusters (indicated by brackets), each involving viruses from two sets of couples, which are shown in greater detail in Fig. 4. One subcluster involved subtype C viruses infecting couples 19 and 61, while the other involved subtype G viruses infecting couples 91 and 148. Given the short genomic region analyzed and the nonsignificant or borderline significant bootstrap values for three of the four couples (couples 19, 61, and 91), we could not determine with confidence that transmission had occurred between the partners of the same rather than different couples. The exact sequence of transmission events involving couples 19 and 61 and couples 91 and 148, respectively, thus remains to be determined. Rapid viral passage from a donor through one or more unidentified intermediaries to his or her putative recipient remains a theoretical possibility for all transmission pairs classified as epidemiologically linked in this study. However, since no other viral subclusters were identified in the data set, the existence of extensive transmission networks within the ZUHRP cohort is highly unlikely.

In summary, this report describes the first comprehensive molecular epidemiological analysis of heterosexual transmission events occurring among discordant couples in an African urban setting. Our analysis allowed us to (i) determine the proportion of linked and unlinked infections with a high level of certainty, (ii) identify the sequence subtype for all transmitted viruses in the genomic regions analyzed, and (iii) examine the cohort for evidence of transmission networks. The results show that of 149 cohabitating couples assumed to have transmitted to each other, 129 (87%) were molecularly confirmed as epidemiologically linked. Nevertheless, approximately 1 in every 10 transmission events involved an individual outside of the partnership. Assumptions concerning transmission linkage based on patient self-reporting alone are thus unlikely to be accurate, and this needs to be factored into the interpretation of transmission data from cohorts in which linkage has not been independently verified. For example, we found a stronger association between plasma viral load and transmission for female-to-male than for male-to-female transmissions in the ZUHRP cohort (12), while such a gender-based difference was not observed in a discordant couple cohort studied in Rakai, Uganda (32). Because transmission linkage was not confirmed at the molecular level, it is possible that some of the putative transmitters in this Ugandan cohort were misclassified. The proportion of unlinked transmissions is likely to vary considerably depending on the demographic, ethnic, and behavioral circumstances characterizing a cohort (45) but will undoubtedly be >0%. Thus, for investigations that require accurate

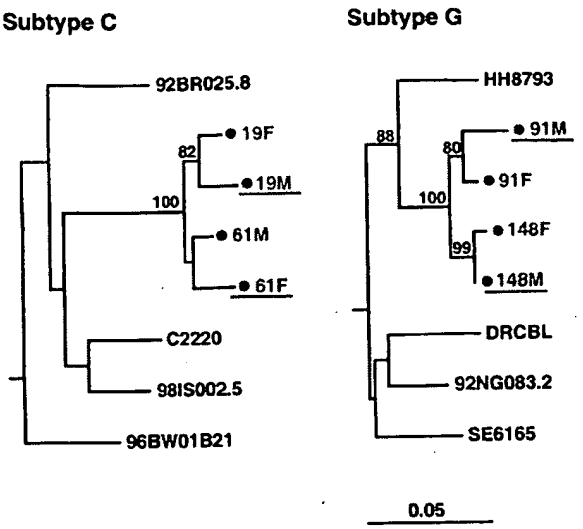


FIG. 4. Transmission networks within the ZUHRP cohort. Phylogenetic trees were constructed from gp41 sequences of viruses infecting four different transmission pairs putatively classified as linked by pairwise sequence analysis. Couple identifiers are indicated in red (F, female partner; M, male partner). Horizontal branch lengths are drawn to scale (the scale bar represents 0.05 nucleotide substitutions per site); vertical separation is for clarity only. Values at nodes indicate the percentage of bootstraps in which the cluster to the right was found; only values of ≥80% are shown. Representative subtype C (left) and subtype G (right) reference sequences are included in each tree. Donor partners are underlined.

FIG. 3. Molecular linkage analysis for a subset of putative HIV-1 transmission pairs. A phylogenetic tree was constructed from partial gp41 sequences (consensus length, 276 bp) by using the neighbor-joining method (34) and the Kimura two-parameter model (22). Horizontal branch lengths are drawn to scale (the scale bar represents 0.05 nucleotide substitutions per site); vertical separation is for clarity only. Asterisks indicate bootstrap values in which the cluster to the right is supported in >80% replicates (out of 1,000). Newly derived sequences from 42 transmission pairs (84 individuals) are shown, along with 26 reference sequences from the Los Alamos Sequence Database (http://hiv-web.lanl.gov/HTML/alignments.html). Viruses from 36 couples are closely related to one another and cluster together with significant bootstrap values, indicating that they are epidemiologically linked (highlighted in red and denoted by dots). Viruses from six couples do not cluster together and exhibit a range of within-couple diversity that is similar to that of the reference sequences (highlighted in blue and denoted by triangles), indicating that they are epidemiologically unlinked. Two small brackets denote viral subclusters, each involving viruses from two sets of couples (see text for details). Brackets on the far right indicate major group M sequence subtypes.

assessment of HIV-1 transmission, such as studies aimed at identifying host and viral transmission correlates or determining the effectiveness of certain prevention strategies, the molecular characterization of viruses from both partners is essential.

We thank the staff, participants, and project management group of the ZUHRP cohort and Maria Salazar for expert technical assistance.

This work was supported by grants N01 AI-85338, R01 AI-40951, and U01 AI-41530 from the National Institutes of Health. DNA sequencing was performed in the DNA Sequence Analysis Core of the UAB Center for AIDS Research, supported by grant P30 A1-27767.

REFERENCES

1. **Albert, J., J. Wahlberg, T. Leitner, D. Escanilla, and M. Uhlen.** 1994. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. J. Virol. **68**:5918–5924.
2. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63–74.
3. **Allen, S., K. E. N'Gandu, and A. Tichacek.** 1998. The evolution of voluntary testing and counseling as an HIV prevention strategy: preventing HIV in developing countries: biomedical and behavioral approaches. Platinum Press, New York, N.Y.
4. **Belec, L., A. Si Mohamed, M. C. Müller-Trutwin, J. Gilquin, L. Gutmann, M. Safar, F. Barre-Sinoussi, and M. D. Kazatchkine.** 1998. Genetically related human immunodeficiency virus type 1 in three adults of a family with no identified risk factor for intrafamilial transmission. J. Virol. **72**:5831–5839.
5. **Blanchard, A., S. Ferris, S. Chamaret, D. Guetard, and L. Montagnier.** 1998. Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. J. Virol. **72**:4537–4540.
6. **Burger, H., B. Weiser, K. Flaherty, J. Gulla, P. N. Nguyen, and R. A. Gibbs.** 1991. Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. Proc. Natl. Acad. Sci. USA **88**:11236–11240.
7. **Carr, J. K., T. Laukkanen, M. O. Salminen, J. Albert, A. Alaeus, B. Kim, E. Sanders-Buell, D. L. Birx, and F. E. McCutchan.** 1999. Characterization of subtype A HIV-1 from Africa by full genome sequencing. AIDS **13**:1819–1826.
8. **Carr, J. K., M. O. Salminen, J. Albert, E. Sanders-Buell, D. Gotte, D. L. Birx, and F. E. McCutchan.** 1998. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology **247**:22–31.
9. **Cassol, S. A., S. Read, B. G. Weniger, P. Gomez, N. Lapointe, C. Y. Ou, and P. G. Babu.** 1996. Dried blood spots collected on filter paper: an international resource for the diagnosis and genetic characterization of human immunodeficiency virus type-1. Mem. Inst. Oswaldo Cruz **91**:351–358.
10. **Faulkner, D. V., and J. Jurka.** 1988. Multiple aligned sequence editor (MASE). Trends Biochem. Sci. **13**:321–322.
11. **Felsenstein, J.** 1992. Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. Genet. Res. **60**: 209–220.
12. **Fideli, U. S., S. A. Allen, R. Musonda, S. Trask, B. H. Hahn, H. Weiss, J. Mulenga, F. Kasolo, S. H. Vermund, and G. M. Aldrovandi.** 2001. Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. AIDS Res. Hum. Retrovir. **17**:901–910.
13. **Frenkel, L. M., J. I. Mullins, G. H. Learn, L. Manns-Arcuino, B. L. Herring, M. L. Kalish, R. W. Steketee, D. M. Thea, J. E. Nichols, S. L. Liu, A. Harmache, X. He, D. Muthui, A. Madan, L. Hood, A. T. Haase, M. Zupancic, K. Staskus, S. Wolinsky, P. Krogstad, J. Zhao, I. Chen, R. Koup, D. Ho, B. Korber, R. J. Apple, R. W. Coombs, S. Pahwa, and N. J. Roberts, Jr.** 1998. Genetic evaluation of suspected cases of transient HIV-1 infection of infants. Science **280**:1073–1077.
14. **Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barre-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1998. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J. Virol. **72**:5680–5698.
15. **Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp.** 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J. Virol. **70**:7013–7029.
16. **Gao, F., L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, B. H. Hahn, and the W.H.O. Network for HIV Isolation and Characterization.** 1994. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. AIDS Res. Hum. Retrovir. **10**:1359–1368.
17. **Gibbons, J. D.** 1997. Inferences concerning location based on two or more samples, p. 169–223. *In* Nonparametric methods for quantitative analysis, 3rd ed. American Sciences Press, Columbus, Ohio.
18. **Goujon, C. P., V. M. Schneider, J. Grofti, J. Montigny, V. Jeantils, P. Astagneau, W. Rozenbaum, F. Lot, C. Frocrain-Herchkovitch, N. Delphin, F. Le Gal, J. C. Nicolas, M. C. Milinkovitch, and P. Deny.** 2000. Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. J. Virol. **74**:2525–2532.
19. **Higgins, D. G., J. D. Thompson, and T. J. Gibson.** 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266**:383–402.
20. **Hillis, D. M., and J. J. Bull.** 1991. Of genes and genomes. Science **254**:528.
21. **Hutchinson, S. J., S. M. Gore, D. J. Goldberg, D. L. Yirrell, J. McGregor, A. G. Bird, and A. J. Leigh-Brown.** 1999. Method used to identify previously undiagnosed infections in the HIV outbreak at Glenochil prison. Epidemiol. Infect. **123**:271–275.
22. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.
23. **Laukkanen, T., J. Albert, K. Liitsola, S. D. Green, J. K. Carr, T. Leitner, F. E. McCutchan, and M. O. Salminen.** 1999. Virtually full-length sequences of HIV type 1 subtype J reference strains. AIDS Res. Hum. Retrovir. **15**: 293–297.
24. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. **73**:152–160.
25. **McKenna, S. L., G. K. Muyinda, D. Roth, M. Mwali, N. Ng'andu, A. Myrick, C. Luo, F. H. Priddy, V. M. Hall, A. A. von Lieven, J. R. Sabatino, K. Mark, and S. A. Allen.** 1997. Rapid HIV testing and counseling for voluntary testing centers in Africa. AIDS **11**:S103–S110.
26. **Novitsky, V. A., M. A. Montano, M. F. McLane, B. Renjifo, F. Vannberg, B. T. Foley, T. P. Ndung'u, M. Rahman, M. J. Makhema, R. Marlink, and M. Essex.** 1999. Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana. J. Virol. **73**:4427–4432.
27. **Oelrichs, R. B., A. M. Vandamme, K. Van Laethem, Z. Debyser, F. E. McCutchan, and N. J. Deacon.** 1999. Full-length genomic sequence of an HIV type 1 subtype G from Kinshasa. AIDS Res. Hum. Retrovir. **15**:585–589.
28. **Oram, J. D., R. G. Downing, M. Roff, J. C. Clegg, D. Serwadda, and J. W. Carswell.** 1990. Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic diversity from other HIV-1 isolates. AIDS Res. Hum. Retrovir. **6**:1073–1078.
29. **Ou, C. Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C. C. Luo, B. T. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, J. Moore, Y. Villamarzo, C. Schable, E. G. Shaper, T. Liberti, S. Lieb, R. Scott, J. Howell, R. Dumbaugh, A. Lasch, B. Kroesen, L. Ryan, K. Bell, V. Munn, D. Marianos, and B. Gooch.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256**:1165–1171.
30. **Piot, P., M. Bartos, P. D. Ghys, N. Walker, and B. Schwartlander.** 2001. The global impact of HIV/AIDS. Nature **410**:968–973.
31. **Poss, M., and J. Overbaugh.** 1999. Variants from the diverse virus population identified at seroconversion of a clade A human immunodeficiency virus type 1-infected woman have distinct biological properties. J. Virol. **73**:5255–5264.
32. **Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, and R. H. Gray.** 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. N. Engl. J. Med. **342**:921–929.
33. **Rodenburg, C. M., Y. Li, S. A. Trask, Y. Chen, J. Decker, D. L. Robertson, M. L. Kalish, G. M. Shaw, S. Allen, B. H. Hahn, and F. Gao.** 2001. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res. Hum. Retrovir. **17**:161–168.
34. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
35. **Salminen, M. O., B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. Leinikki, D. S. Burke, and F. E. McCutchan.** 1996. Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. AIDS Res. Hum. Retrovir. **12**:1329–1339.
36. **Song, J. Z., B. Wang, Y. C. Ge, D. E. Dwyer, A. L. Cunningham, and N. K. Saksena.** 1999. Significance of plasma and peripheral blood mononuclear cell derived HIV-1 sequences in establishing epidemiologic linkage between two individuals multiply exposed to HIV-1. Microb. Pathog. **26**:287–298.
37. **Spire, B., J. Sire, V. Zachar, F. Rey, F. Barre-Sinoussi, F. Galibert, A. Hampe, and J. C. Chermann.** 1989. Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. Gene **81**:275–284.
38. **Srinivasan, A., R. Anand, D. York, P. Ranganathan, P. Feorino, G. Schochetman, J. Curran, V. S. Kalyanaraman, P. A. Luciw, and R. Sanchez-Pescador.**

1987. Molecular characterization of human immunodeficiency virus from Zaire: nucleotide sequence analysis identifies conserved and variable domains in the envelope gene. Gene 52:71–82.
39. UNAIDS-W.H.O. 2000. AIDS epidemic update: December 2000. Joint United Nations Programme on HIV/AIDS and World Health Organization. World Health Organization, Geneva, Switzerland.
40. UNAIDS-W.H.O. 2000. AIDS in Africa, country by country. Joint United Nations Programme on HIV/AIDS and World Health Organization. World Health Organization, Geneva, Switzerland.
41. **Vartanian, J. P., A. Meyerhans, B. Asjo, and S. Wain-Hobson.** 1991. Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. 65:1779–1788.
42. **Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology 189:103–110.
43. **Yang, C., D. Pieniazek, S. M. Owen, C. Fridlund, J. Nkengasong, T. D. Mastro, M. A. Rayfield, R. Downing, B. Biryawaho, A. Tanuri, L. Zekeng, G. van der Groen, F. Gao, and R. B. Lal.** 1999. Detection of phylogenetically diverse human immunodeficiency virus type 1 groups M and O from plasma by using highly sensitive and specific generic primers. J. Clin. Microbiol. 37:2581–2586.
44. **Yirrell, D. L., S. J. Hutchinson, M. Griffin, S. M. Gore, A. J. Leigh-Brown, and D. J. Goldberg.** 1999. Completing the molecular investigation into the HIV outbreak at Glenochil prison. Epidemiol. Infect. 123:277–282.
45. **Yirrell, D. L., H. Pickering, G. Palmarini, L. Hamilton, A. Rutemberwa, B. Biryahwaho, J. Whitworth, and A. J. Leigh Brown.** 1998. Molecular epidemiological analysis of HIV in sexual networks in Uganda. AIDS 12:285–290.

JOURNAL OF VIROLOGY, June 2001, p. 5526–5540
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.12.5526–5540.2001

Vol. 75, No. 12

The Ability of an Oligomeric Human Immunodeficiency Virus Type 1 (HIV-1) Envelope Antigen To Elicit Neutralizing Antibodies against Primary HIV-1 Isolates Is Improved following Partial Deletion of the Second Hypervariable Region

S. W. BARNETT,[1] S. LU,[2] I. SRIVASTAVA,[1] S. CHERPELIS,[3] A. GETTIE,[1,4] J. BLANCHARD,[4] S. WANG,[2] I. MBOUDJEKA,[2] L. LEUNG,[1] Y. LIAN,[1] A. FONG,[1] C. BUCKNER,[3] A. LY,[3] S. HILT,[1] J. ULMER,[1] C. T. WILD,[5] J. R. MASCOLA,[6]† AND L. STAMATATOS[3]*

Chiron Corporation, Emeryville, California 94608-2916[1]; University of Massachusetts Medical Center, Worcester, Massachusetts 01655[2]; Aaron Diamond AIDS Research Center, The Rockfeller University, New York, New York 10016[3]; Tulane Regional Primate Research Center, Covington, Louisiana 70433[4]; Panacos Pharmaceuticals, Gaithersburg, Maryland 20877[5]; and Walter Reed Army Institute of Research, Rockville, Maryland 20850[6]

Received 6 February 2001/Accepted 20 March 2001

Partial deletion of the second hypervariable region from the envelope of the primary-like SF162 virus increases the exposure of certain neutralization epitopes and renders the virus, SF162ΔV2, highly susceptible to neutralization by clade B and non-clade B human immunodeficiency virus (HIV-positive) sera (L. Stamatatos and C. Cheng-Mayer, J. Virol. 78:7840–7845, 1998). This observation led us to propose that the modified, SF162ΔV2-derived envelope may elicit higher titers of cross-reactive neutralizing antibodies than the unmodified SF162-derived envelope. To test this hypothesis, we immunized rabbits and rhesus macaques with the gp140 form of these two envelopes. In rabbits, both immunogens elicited similar titers of binding antibodies but the modified immunogen was more effective in eliciting neutralizing antibodies, not only against the SF162ΔV2 and SF162 viruses but also against several heterologous primary HIV type 1 (HIV-1) isolates. In rhesus macaques both immunogens elicited potent binding antibodies, but again the modified immunogen was more effective in eliciting the generation of neutralizing antibodies against the SF162ΔV2 and SF162 viruses. Antibodies capable of neutralizing several, but not all, heterologous primary HIV-1 isolates tested were elicited only in macaques immunized with the modified immunogen. The efficiency of neutralization of these heterologous isolates was lower than that recorded against the SF162 isolate. Our results strongly suggest that although soluble oligomeric envelope subunit vaccines may elicit neutralizing antibody responses against heterologous primary HIV-1 isolates, these responses will not be broad and potent unless specific modifications are introduced to increase the exposure of conserved neutralization epitopes.

Analysis of the crystal structure of the gp120 human immunodeficiency virus (HIV) envelope subunit indicated that neutralization epitopes are primarily clustered in one face of this protein, which is naturally occluded within the oligomeric envelope form, i.e., that present on the surface of virions and infected cells (16, 37). These structural observations are supported by numerous immunochemical and virological studies (1, 24, 25, 27, 28, 31, 35, 38, 40).

Several reports have indicated that specific modifications (such as deglycosylations and loop deletions) introduced in the envelope glycoproteins of HIV and simian immunodeficiency virus (SIV) may increase the exposure of neutralization epitopes. Wyatt et al. demonstrated that on the background of the HXB2 virus, a laboratory-adapted CXCR4-using (X4-using) virus, deletions of the first, second, and third hypervariable regions (V1, V2, and V3 loops, respectively) of the gp120 envelope subunit increase the exposure of epitopes participating in HIV envelope-CD4 and -coreceptor binding (38, 40). Subsequently, it was demonstrated that the simultaneous deletion of the V1 and V2 loops from the envelope of this virus increases it susceptibility to neutralization by anti-V3 loop and certain CD4-induced monoclonal antibodies (MAbs) (3). Reitter et al. reported that elimination of specific asparagine-linked glycosylation sites located in the V1 loop of SIVmac239 results in the exposure of neutralization epitopes and, importantly, increases their immunogenicity (25). Infection of macaques with SIVmac239-derived viruses expressing such partially deglycosylated envelopes results in the generation of antienvelope antibodies capable of neutralizing the parental virus SIVmac239, which displays a fully glycosylated envelope, more efficiently than antibodies elicited during infection of macaques with SIVmac239 itself.

We previously reported that on the background of the SF162 virus, a primary-like CCR5-using (R5-using) isolate, deletion of the 30 amino acids from the central region of the V2 loop (SF162ΔV2) does not abrogate its infectivity but renders it highly susceptible to neutralization by sera collected from patients infected with heterologous HIV type 1 (HIV-1) isolates (30). We hypothesized that on the background of the SF162 envelope, partial elimination of the V2 loop increases the

* Corresponding author. Mailing address: Aaron Diamond AIDS Research Center, 455 First Ave., 7th Floor, New York, NY 10016. Phone: (212) 448-5000. Fax: (212) 725-1126. E-mail: lstamata@adarc.org.

† Present address: Vaccine Research Center, National Institutes of Health, Bethesda, MD 20892.

exposure of neutralization epitopes that are conserved among heterologous primary HIV-1 isolates.

In this study, we compared the immunogenic potentials of the unmodified SF162 and modified SF162ΔV2 (hereafter designated ΔV2) envelopes. Using the gene gun vaccination method, we immunized rabbits with the gp140 form of the SF162 and ΔV2 envelopes. We observed that both immunogens elicited the generation of similar antibody titers, but that the modified immunogen elicited higher titers of neutralizing antibodies against the parental SF162 virus than the unmodified immunogen. These results are in agreement with those previously reported in the case of SIVmac239 (25), because they suggest that specifically modified envelope immunogens are more effective than the corresponding unmodified envelope immunogens in eliciting neutralizing antibodies against the homologous parental virus. Additionally, the ΔV2-derived modified immunogen was more effective than the SF162-derived unmodified immunogen in generating antibodies capable of neutralizing heterologous primary HIV-1 isolates.

The immunogenicity of these two antigens was also evaluated in rhesus macaques, an animal model more closely related to humans and more suitable for HIV vaccine studies, using the DNA-prime–protein-boost vaccination method. Here too we recorded that the modified immunogen was more effective than the unmodified immunogen in generating potent neutralizing antibodies both against the homologous SF162ΔV2 and parental SF162 viruses. The antibodies elicited in macaques by the modified, but not unmodified, immunogen neutralized several, but not all, heterologous primary HIV-1 isolates. The neutralizing potential against the heterologous isolates tested was lower than that against the parental SF162 virus. Previous studies reported that cross-reactive neutralizing antibodies against primary HIV-1 isolates could be elicited in mice immunized with fusion-competent vaccines (18) or soluble oligomeric envelopes derived from a primary-like HIV envelope (41). Our studies indicate for the first time that potent cross-reactive neutralizing antibodies can be elicited in nonhuman primates immunized with soluble oligomeric subunit HIV envelope vaccines derived from an R5-using primary-like HIV-1 isolate. They strongly suggest, however, that specific envelope modifications can be introduced to increase the exposure of neutralization epitopes and increase the breadth and potency of these responses.

MATERIALS AND METHODS

Viruses. The isolation and phenotypic characterization of the SF162 and SF162ΔV2 isolates were previously reported (5, 30). The primary clade B HIV-1 isolates 92US660, 92HT593, 92US657, 92US714, 92US727, 91US056, 91US054 and 93US073 were obtained from the NIH AIDS Research and Reference Reagent Program. All viral stocks were prepared and titrated in activated human peripheral blood mononuclear cells (PBMC).

Vaccines. The DNA vector used to express our immunogens in rabbits is the pJW4303 (20). The DNA vector used to immunize rhesus macaques is derived from pCMVKm2 (4, 43). Both DNA plasmids contain the human cytomegalovirus enhancer/promoter elements, and the native leader peptide of the HIV envelope was replaced with that derived from the tissue-specific plasminogen activator gene. In the case of macaque immunizations, the DNA construct was codon optimized for high expression in mammalian cells. Both DNA vectors express the gp140 ectodomain form of the HIV envelope immunogen, with an intact gp120-gp41 cleavage site.

Protein-boosting immunizations were performed only in rhesus macaques to increase the titer of antibodies elicited following the DNA phase of immunization. For this purpose, the ΔV2 gp140 protein was produced in CHO cells and purified as stable soluble trimers (I. Srivastava et al., unpublished data). To increase the stability of these secreted oligomers, the gp120-gp41 cleavage site was eliminated by mutagenesis (9, 10, 32).

Immunizations (i) Rabbits. Each animals received five DNA immunizations (each immunization consisting of 36 shots of 0.5 μg of DNA each) by the gene gun vaccination method (20) at weeks 0, 4, 8, 18, and 22. Blood was drawn 2 weeks following each immunization. Six animals (A1 to A6) were immunized with the unmodified SF162 gp140 immunogen, and six animals (A7 to A12) received the modified ΔV2 gp140 immunogen. Two animals (A13 and A14) served as controls and were immunized with the DNA vector alone.

(ii) Rhesus macaques. Animals H445 and J408 were immunized with the modified ΔV2 gp140 immunogen, animals N472 and P655 were immunized with the unmodified SF162 gp140 immunogen, and animals M844 and H473 were immunized with the DNA vector alone. Before immunization, the animals were tested for antibodies to various simian viruses such as SIV, type D retroviruses, and simian T-lymphocytic virus type 1. Animals vaccinated with the modified envelope were immunized with DNA at weeks 0, 4, and 8, and animals vaccinated with the unmodified envelope were immunized with DNA at weeks 0, 4, and 9. The DNA (2 mg of DNA in 1 ml of endotoxin-free water each time per animal) was administered both intradermally at two sites (0.2 mg at each site) and intramuscularly (0.8 mg at two sites in the quadriceps muscles). Animals were immunized a fourth time with DNA and at the same time with the purified oligomeric ΔV2 or SF162 gp140 protein mixed with the adjuvant MF-59C. The proteins (0.1 mg of purified protein in 0.5 ml [total volume] per animal) were administered intramuscularly in the deltoids. The control animals received only adjuvant. This DNA-plus-protein booster immunization took place at week 27 for animals vaccinated with the modified immunogen and at week 48 for animals immunized with the unmodified immunogen. At week 38, the animals immunized with the modified, but not those immunized with the unmodified, immunogen were immunized one additional time with the adjuvanted protein alone (no DNA).

Antibody determination (i) Anti-gp140 antibodies. Titers were determined throughout the immunization protocol by enzyme-linked immunosorbent assay (ELISA) as previously described (31, 33). Briefly, purified soluble oligomeric ΔV2 and SF162 gp140 proteins were used to coat ELISA plates (Immulon 2HB) (0.2 μg of protein in 0.1 ml of 100 mM $NaHCO_3$ [pH 8.5]) by overnight incubation at 4°C. Nonadsorbed protein molecules were removed by washing with Tris-buffered saline (TBS), and the wells were blocked with SuperBlock (SB; Pierce). Heat-inactivated (56°C for 35 min) sera collected from the immunized animals were serially diluted in SB and added to the wells (0.1 ml per well) for 1 h at 37°C. In the case of rabbits, sera from control animals receiving the DNA vector alone were used as negative controls. In the case of macaques, preimmunization sera were used as negative controls. Unbound antibodies were removed by TBS washing, and the envelope-bound antibodies were detected with the use of goat anti-human (in the case of rhesus sera) or anti-rabbit (in the case of rabbit sera) immunoglobulin G coupled to alkaline phosphatase antibodies (Zymed Immunochemicals) as previously described (31). The optical density at 490 nm (OD_{490}) of each well was recorded with a Bioluminometer (Molecular Dynamics). A plot of the OD_{490} signals versus serum dilution was generated, and endpoint antibody titers were determined as the highest postimmunization serum dilution that produces an OD_{490} value three times that of the OD_{490} produced by the preimmunization sera at their lowest dilution. Sera from various stages of immunization were tested at the same time.

(ii) Anti-V3 loop antibodies. Titers of anti-V3 loop antibodies generated during immunization were determined by ELISA using the peptide CKSITIGPG RAFYATGDC, derived from the central region of the SF162-derived V3 loop. This peptide was diluted in 0.2 M sodium bicarbonate (pH 9.4) at a concentration of 1 μg/ml and then used to coat ELISA plates (0.1 ml per well) by overnight incubation at 37°C. The wells were washed with TBS and blocked with SB as described above. Sera collected before and after immunization were serially diluted in SB containing 0.3% (vol/vol), Tween and added (0.1 ml per well) for 2 h at room temperature. The plates were washed with TBS containing 0.3% (vol/vol) Tween, and V3 loop-bound antibodies were detected as described above with the use of immunoglobulin G coupled to alkaline phosphatase antibodies of the appropriate species. The antibody titers were determined as described above.

Neutralization assays. Neutralization assays were performed using as target cells human PBMC activated for 3 days with phytohemagglutinin (3 μg/ml; Sigma) as previously described (21, 22, 30, 34). All HIV-1 isolates tested were grown and titrated in human PBMC, aliquoted, and kept frozen at −80°C until further use. Viruses (50 to 100 50% tissue culture infective doses in 50 μl of complete RPMI medium containing 20 U of interleukin-2 [Hoffmann-La Roche] per ml) were preincubated with an equal volume of serially diluted heat-inacti-

vated (35 min at 56°C) sera for 1 h at 37°C in 96-well U-bottom plates (Corning). For each serum dilution, triplicate wells were used. Preimmunization sera from macaques and sera collected from rabbits immunized with the DNA vector alone were also incubated with the viruses and served as controls for nonspecific neutralization. To each well, 0.1 ml of complete medium containing 0.4×10^6 phytohemagglutinin-activated PBMC was added. Following overnight incubation at 37°C, half of the volume of each well was replaced with fresh, complete RPMI medium. Following centrifugation of the plates (5 min at 2,000 rpm), half of the volume of each well was again replaced with fresh medium. This procedure was repeated twice. The p24 antigen concentration in each well was evaluated at various points following infection (usually at days 4, 6, and 11), using an in-house ELISA p24- detection assay. The mean percent neutralization from triplicate wells and the standard deviation for each serum dilution were calculated based on p24 concentrations recorded in wells containing virus, cells, and no rabbit or macaque serum as previously described (34). However, we noticed that infection of some isolates was reduced in the presence of preimmunization sera (nonspecific neutralization). We decided therefore to present the results from our neutralization studies in two ways: (i) in the same figure, we present both the neutralization curve recorded with sera collected prior to vaccination (prebleeds) and that recorded with sera collected at various stages following vaccination; (ii) for each serum dilution, we calculate the difference between the percent neutralization recorded with postvaccination sera minus that recorded with prevaccination sera. In some figures, this difference (which we term specific neutralization) is plotted as a function of serum dilution. In parallel, we evaluated the susceptibilities of the various primary isolates to neutralization by MAbs 2F5 and 2G12.

During these neutralization experiments, we also evaluated the abilities of sera collected from macaques immunized with the recombinant SF2 gp120 envelope. This immunogen was previously tested as a potential vaccine against HIV and failed to raise cross-reactive neutralizing antibodies (22).

RESULTS

Generation of antibodies in rabbits. Both the SF162- and SF162ΔV2-derived immunogens elicited high titers of antibodies capable of binding to both oligomeric ΔV2 and SF162 gp140 (Fig. 1). As expected, variantions in the antibody titers were recorded throughout the vaccination schedule in animals belonging to either group. However, no statistically significant differences in antibody titers were recorded between the two animal groups throughout the immunization schedule. The antibody titers in each animal, regardless of whether it was immunized with the modified or the unmodified immunogen, were very weak during the first two immunizations (at 0 and 4 weeks). The fourth immunization (at 18 weeks) resulted in an increase in antibody titers, compared to the third immunization (8 weeks), between 2 and 3 $\log_{10}$ in both animal groups. The fifth immunization (22 weeks) increased the antibody titers, compared to the fourth immunization, against the SF162 gp140 antigen (by less than 1 $\log_{10}$) but not against the ΔV2 gp140 protein. At the end of the vaccination schedule, very potent endpoint ELISA binding antibody titers in the order of 10^5 to 10^6 were recorded in both animal groups against both antigens. Thus, it appears that in rabbits, based on the assay used here to determine antibody titers, the modified immunogen is as effective as the unmodified immunogen in eliciting the generation of antibodies even though the former immunogen lacks 30 amino acids from the V2 loop.

Neutralizing activity in rabbit sera against the SF162 and SF162ΔV2 isolates. Both immunogens generated neutralizing antibodies against the SF162ΔV2 virus following the third DNA immunization (Fig. 2A). A trend toward higher neutralization titers in the modified immunogen-vaccinated group was recorded. Thus, the mean (± standard error) serum dilutions at which 70% inhibition of infection was recorded for SF162 gp140- and ΔV2 gp140-immunized animals were 179 (±34) and 483 (±148), respectively. At this stage of vaccination, while two (A8 and A9) out of six animals immunized with the modified immunogen elicited neutralizing antibodies against the parental SF162 isolate, none of the animals immunized with the unmodified immunogen elicited antibodies capable of doing so (Fig. 2B). However, the number of animals that generated neutralizing antibodies against the SF162 and SF162ΔV2 viruses increased with each subsequent immunization, so that at the end of the immunization schedule (i.e., after the fifth immunization) all animals had generated neutralizing antibodies against the SF162 virus. In addition, the neutralizing potency of each serum, regardless of whether the animal was vaccinated with the modified or unmodified immunogen, increased with each immunization.

At the end of the immunization schedule, sera collected from rabbits immunized with the modified immunogen had higher neutralization potency against both SF162ΔV2 and SF162 viruses than sera collected from animals immunized with the unmodified immunogen. Six out of six animals immunized with the modified immunogen elicited antibodies capable of neutralizing the SF162ΔV2 virus between 70 and 100% at a 1:5,000 dilution (Fig. 2A). In contrast, at the same serum dilution only one (A1) of the six animals vaccinated with the unmodified envelope developed antibody responses able to neutralize SF162ΔV2 infection, and that by only 50%. The remaining five animals in this group failed to elicit antibody responses potent enough to neutralize SF162ΔV2 infection to any significant extent at this dilution. Differences in neutralizing potential between sera collected from animals immunized with the modified immunogen and those immunized with the unmodified immunogen were also evident when their abilities to neutralize the SF162 virus were compared (Fig. 2B). Sera collected from four (A8, A9, A10, and A12) out of six animals immunized with the modified antigen neutralized SF162 infection between 70 and 90% at 1:100 to 1:300 dilutions. In contrast, none of the sera collected from animals immunized with the unmodified antigen could inhibit SF162 infection by 70 to 90% at the same dilutions.

Generation of cross-reactive neutralizing antibodies in rabbits. The fact that the SF162ΔV2-derived envelope immunogen was capable of eliciting higher titers of neutralizing antibodies against the parental SF162 isolate (which expresses the full envelope) than the immunogen derived from the SF162 isolate itself prompted us to examine whether the modified immunogen was also more effective in eliciting cross-reactive neutralizing antibodies, i.e., antibodies capable of neutralizing heterologous to the vaccine primary HIV-1 isolates. We tested several such isolates whose neutralization susceptibility to various MAbs was previously documented (8). Only two (92US714 and the 92HT593) out of the six isolates examined were neutralized by antibodies elicited by the unmodified immunogen (Table 1). All animals except animal A1 developed neutralizing antibodies against 92US714, while only animals A2 and A5 generated neutralizing antibodies against 92HT593. In contrast, four out of the six animals immunized with the modified ΔV2 gp120 immunogen generated cross-reactive neutralizing antibodies against most of the heterologous isolates tested. In addition, the neutralization potency of sera collected from animals immunized with the modified immunogen was

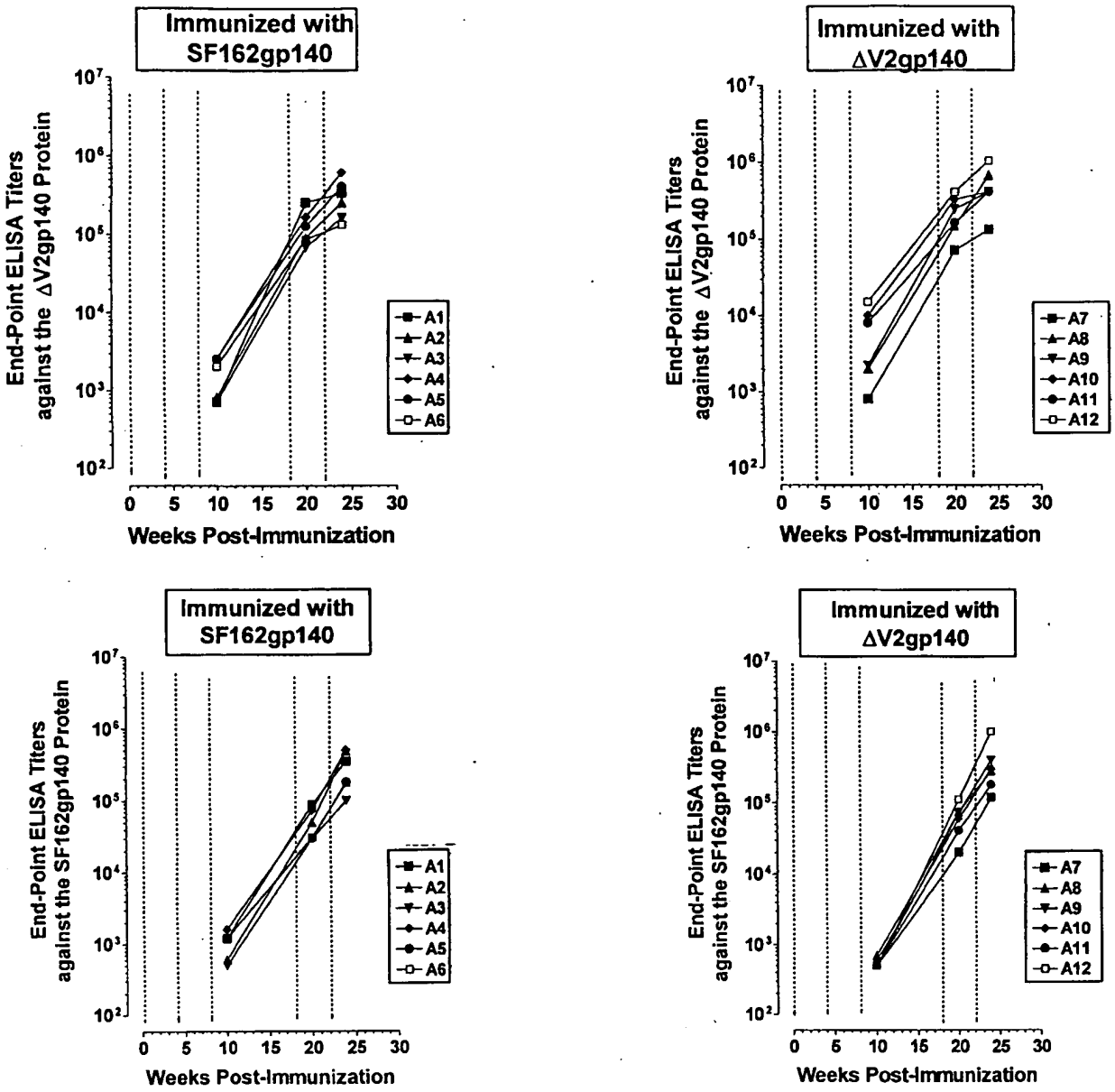


FIG. 1. Development of antibodies in rabbits. Six animals (A1 to A6) were immunized with DNA expressing the unmodified SF162 gp140 immunogen, and six (A7 to A12) were immunized with DNA expressing the modified ΔV2 gp140 immunogen. Titers were determined 2 weeks following each immunization, as described in Materials and Methods, against the oligomeric SF162 and ΔV2 gp140 proteins. Dashed lines indicate times of immunizations.

higher than that of sera collected from animals immunized with the unmodified immunogen (Table 1). Thus, although 80% inhibition of infection was frequently recorded with the former sera, this level of inhibition was recorded in only two instances (sera from animal A5 versus the 92US714 and 92HT593 isolates).

Development of antibodies in rhesus macaques vaccinated with the modified ΔV2 gp140 immunogen. The above results prompted us to evaluate the immunogenic potential of the unmodified SF162 gp140 and modified ΔV2 gp140 antigens in rhesus macaques, an animal model where the protective potential of vaccine-elicited antibodies can eventually be evaluated. Macaques were vaccinated with these two immunogens by the DNA-prime–protein-boost vaccination method.

Envelope-specific antibodies became detectable following the second DNA immunization (Fig. 3). At this stage, endpoint ELISA titers in animals immunized with the modified antigen (animals J408 and H445) were in the order of 1:2,000. In contrast, in animals immunized with the unmodified envelope (animals N472 and P655), antibodies were detectable only in animal N472 (endpoint ELISA titers in the order of 1:500). With the exception of animal H445, the third DNA immuni-

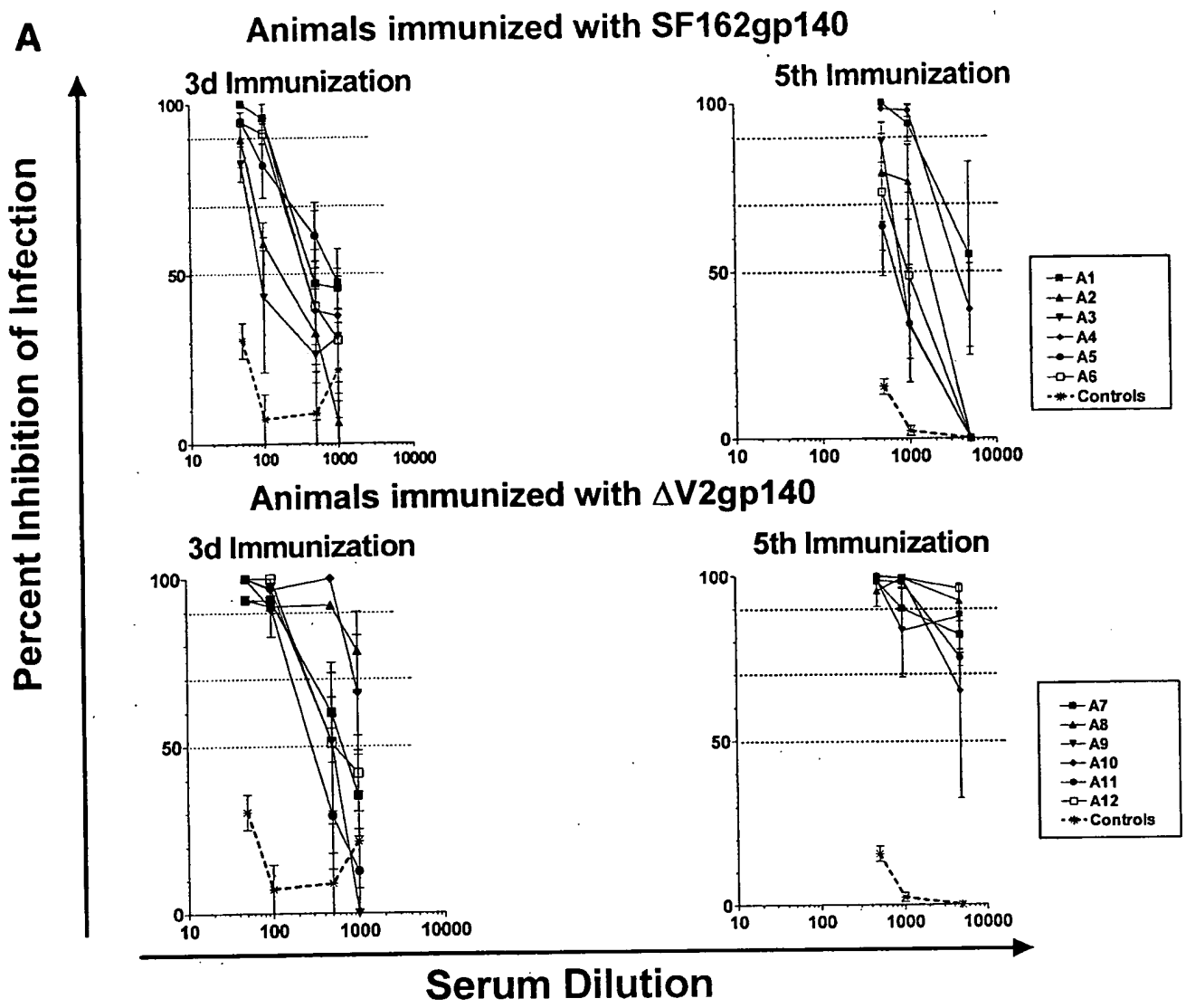


FIG. 2. Results of neutralization experiments using rabbit sera collected following the third and fifth immunizations against the SF162ΔV2 (A) and SF162 (B) viruses. Data are representative of at least three independent experiments. Symbols indicate the mean percent neutralization and standard deviation from triplicate wells. Dashed lines indicate 50, 70, and 90% inhibition of infection. Asterisks (controls) represent neutralization curves obtained with sera collected from animals that were immunized with the DNA vector alone and are indicative of nonspecific neutralization.

zation did not further increase the antibody titers. Anti-gp120 and anti-gp41 antibodies were generated synchronously during DNA immunization (data not shown).

During the subsequent 5 to 10 months of observation, antibodies were undetectable in animals immunized with the unmodified SF162 gp140 immunogen, while in animals immunized with the modified ΔV2 gp140 immunogen the antibodies were always detectable, but their titers declined over time.

Following the DNA-plus-protein booster immunization, the antibody titers increased significantly in all animals. At their peak values (reached within 2 to 4 weeks postboosting), endpoint ELISA antibody titers in animals immunized with the modified ΔV2 gp140 immunogen were 1:30,000 for animal J408 and 1:110,000 for animal H445. The titers decreased gradually over time and remained stable at approximately 1:8,000 for several weeks in both animals. Higher peak antibody titers were recorded in animals vaccinated with the unmodified SF162 gp140 immunogen (endpoint ELISA antibody titers of 1:150,000 in animal N472 and 175,000 in animal P655). During the following 7 weeks of observation, the antibody titers decreased more rapidly in both animals to approximately 1:35,000. Thus, in contrast to what we recorded in rabbits, in macaques the unmodified immunogen generated higher titers of binding antibodies than the modified immunogen.

As expected, anti-HIV envelope antibodies were not generated in control animals (M844 and H473) immunized with the DNA vector alone.

Neutralizing activity of macaque sera against the homologous SF162ΔV2 and parental SF162 isolates. During the DNA phase of immunization, only animals immunized with the mod-

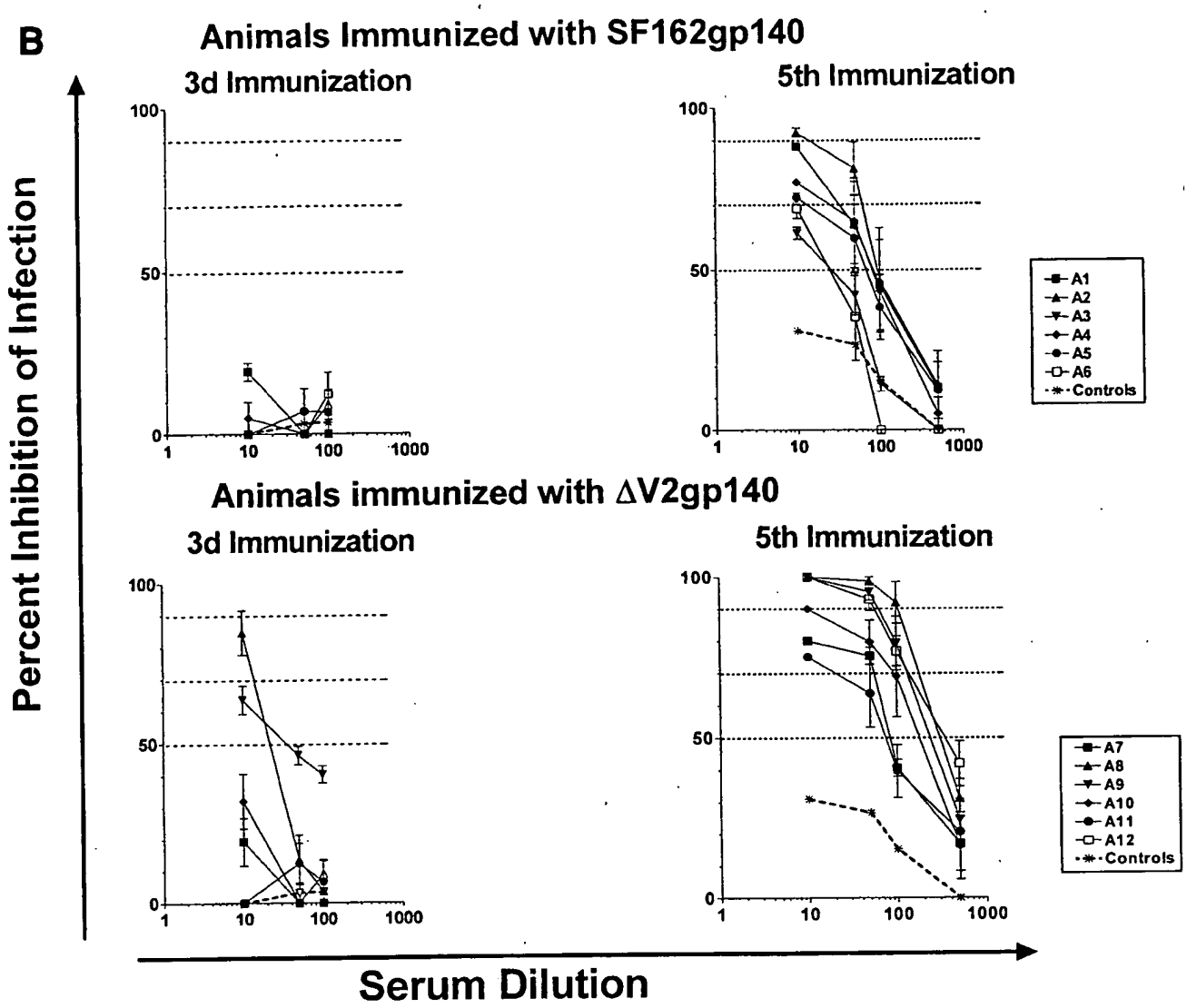


FIG. 2—*Continued.*

ified ΔV2 gp140 immunogen elicited neutralizing antibodies against the SF162 and SF162ΔV2 viruses (Fig. 4). Following the second DNA immunization, animal J408 developed neutralizing antibodies against the homologous SF162ΔV2, but not the parental SF162, isolate (Fig. 4A). The titer of neutralizing antibodies in animal J408 increased following the third DNA immunization, at which point neutralization of both isolates was recorded, although the titers of binding antibodies did not increase in parallel (Fig. 3). In contrast, much weaker neutralizing antibody responses against the SF162ΔV2 and no neutralizing responses against the SF162 virus were elicited in animal H445, even though this animal generated titers of binding antibodies similar to those generated in animal J408 (Fig. 3).

Two weeks following the DNA-plus-protein booster immunization, sera collected from animals immunized with either immunogen inhibited SF162ΔV2 infection. The neutralization potency of sera collected from animals immunized with the modified immunogen was higher than that of sera collected from animals immunized with the unmodified immunogen. For example, 50% inhibition of SF162ΔV2 infection was recorded at dilutions of 1:2,000 to 1:5,000 from the former sera, but this level of inhibition was not recorded at this dilutions with the latter sera. Both ΔV2 gp140-immunized animals generated strong neutralizing antibodies against the parental SF162 virus, while only one (N472) of the two animals immunized with the SF162 gp140 immunogen generated neutralizing antibodies against this virus. Changes in the neutralizing potency of these sera were not recorded during the subsequent 2 weeks, even tough changes in the antibody titer levels were detectable during this period (Fig. 3). Control animals (M844 and H473) vaccinated with the vector alone did not develop neutralizing antibodies (data not shown).

Neutralization of heterologous primary HIV-1 isolates by macaques sera. The breath of the neutralizing antibody responses elicited in macaques immunized with the modified and unmodified immunogens was evaluated by comparing the abil-

TABLE 1. Generation of cross-reactive neutralizing antibodies in rabbits

gp140	Animal	Specific neutralization[a]											
		91US054		92US657		92US660		92HT593		91US056		92US714	
		50%	80%	50%	80%	50%	80%	50%	80%	50%	80%	50%	80%
Unmodified SF162	A1	−	−	−	−	−	−	−	−	−	−	−	−
	A2	−	−	−	−	−	−	+	−	−	−	+	−
	A3	−	−	−	−	−	−	−	−	−	−	+	−
	A4	−	−	−	−	−	−	−	−	−	−	+	−
	A5	−	−	−	−	−	−	+	+	−	−	+	+
	A6	−	−	−	−	−	−	−	−	−	−	+	−
Modified ΔV2	A7	+	+	−	−	−	−	+	+	+	+	+	−
	A8	+	+	−	−	−	−	+	+	+	+	+	+
	A9	+	−	+	−	−	−	+	+	+	+	+	+
	A10	−	−	+	+	−	−	+	+	−	−	+	+
	A11	−	−	−	−	−	−	−	−	−	−	−	−
	A12	−	−	−	−	−	−	−	−	−	−	−	−

[a] Neutralizing activity was evaluated at 1:10 dilution, taking into consideration the nonspecific neutralization recorded with sera collected from animals vaccinated with the DNA vector alone (see Materials and Methods for details). −, 50% specific neutralization was not recorded; +, 50 or 80% specific neutralization was recorded. Results are from three independent neutralization experiments.

ities of sera collected from macaques immunized with these two immunogens to block infection of heterologous primary clade B HIV-1 isolates. During our serum neutralization experiments, we evaluated in parallel the susceptibility of these isolates to neutralization by two of the most commonly used primary-isolate-neutralizing MAbs (2F5 and 2G12) (Table 2).

Heterologous isolate neutralization was not recorded (less than 50% inhibition of infection at 1:10 serum dilution) during the DNA phase of immunization in macaques (data not shown). Two weeks following the DNA-plus-protein booster immunization, sera collected from the two animals vaccinated with the modified ΔV2 gp140 protein neutralized some of the heterologous primary HIV-1 isolates tested (Fig. 5). At the lowest serum dilution tested (1:10), and when nonspecific neutralization recorded with preimmunization sera was taken into consideration (see Materials and Methods for details), 80 to 90% inhibition of infection was recorded only with the ADA, 91US056, and 92US714 isolates by J408 sera and with the ADA, 92US714, and 92US660 isolates with the H445 sera (Fig. 5 and Table 2). The cross-neutralizing activity of the sera collected from these two animals differed. For example, 92US660 infection was inhibited by 80 and 50% by H445 and J408 sera, respectively. The serum cross-neutralizing activity decreased during the subsequent weeks of observation (Fig. 5). Sera collected 5 weeks following this DNA-plus-protein booster immunization had no cross-reactive neutralizing activity, even though potent neutralization of the SF162 and SF162ΔV2 isolates was still recorded.

Despite the fact that following this DNA-plus-protein booster immunization, the binding antibody titers in animals vaccinated with the unmodified immunogen were higher than those in animals vaccinated with the modified immunogen (Fig. 3), the former sera failed to neutralize any of the heterologous isolates tested (Table 2) (i.e., less than 50% specific neutralization was recorded). Thus, although in rabbits the unmodified immunogen was able to elicit (albeit much less efficiently than the modified immunogen) neutralizing antibodies against some heterologous primary HIV-1 isolates (Table 1), it failed to do so in rhesus macaques.

In parallel, we evaluated the susceptibility of the heterologous isolates to neutralization by sera collected from macaques that had been immunized with the recombinant SF2-derived gp120 protein. This protein was previously evaluated as a vaccine candidate and was ineffective in eliciting cross-reactive neutralizing antibodies; i.e., less than 50% neutralization at serum dilutions of 1:10 was recorded (22). All of the isolates tested here were not susceptible to neutralization by antibodies elicited by the SF2 gp120 protein (Table 2).

Second booster immunization with the modified ΔV2 gp140 protein. Although the above results indicated that the modified ΔV2 gp140 immunogen was indeed more effective in eliciting cross-reactive neutralizing antibody responses than the unmodified immunogen, these responses were weaker than those recorded against the parental SF162 isolate (Fig. 5). In an effort to further increase the potency and breath of these responses, we attempted to further boost the antibody titers in animals H445 and J408 by immunizing them one additional time with the purified oligomeric ΔV2 gp140 protein (this time in the absence of DNA immunization).

An increase in antibody titers was indeed recorded following this protein boost, so that at their peak values (1:145,000 endpoint ELISA titers) the titers were approximately threefold higher than those recorded during the first booster immunization with DNA plus protein (Fig. 6A). In parallel, we recorded a significant increase in the titer of neutralizing antibodies against the homologous SF162ΔV2 and parental SF162 isolates (Fig. 6B). No differences in the neutralizing potential of the sera collected 2 and 5 weeks following this last boost were recorded, even though the binding antibody titers decreased significantly during the same period. Unexpectedly, however, the neutralizing potential of the same sera against most of the heterologous primary isolates tested generally decreased (Table 2). Thus, with the exception of the BZ167, 92US657, and ADA isolates, all of the heterologous isolates tested were resistant to neutralization by sera collected 2 weeks following the second boost. Interestingly, although isolate 92US657 was resistant to neutralization by sera collected following the first boost, it became susceptible to neutralization by sera collected from animal H445 following the second boost.

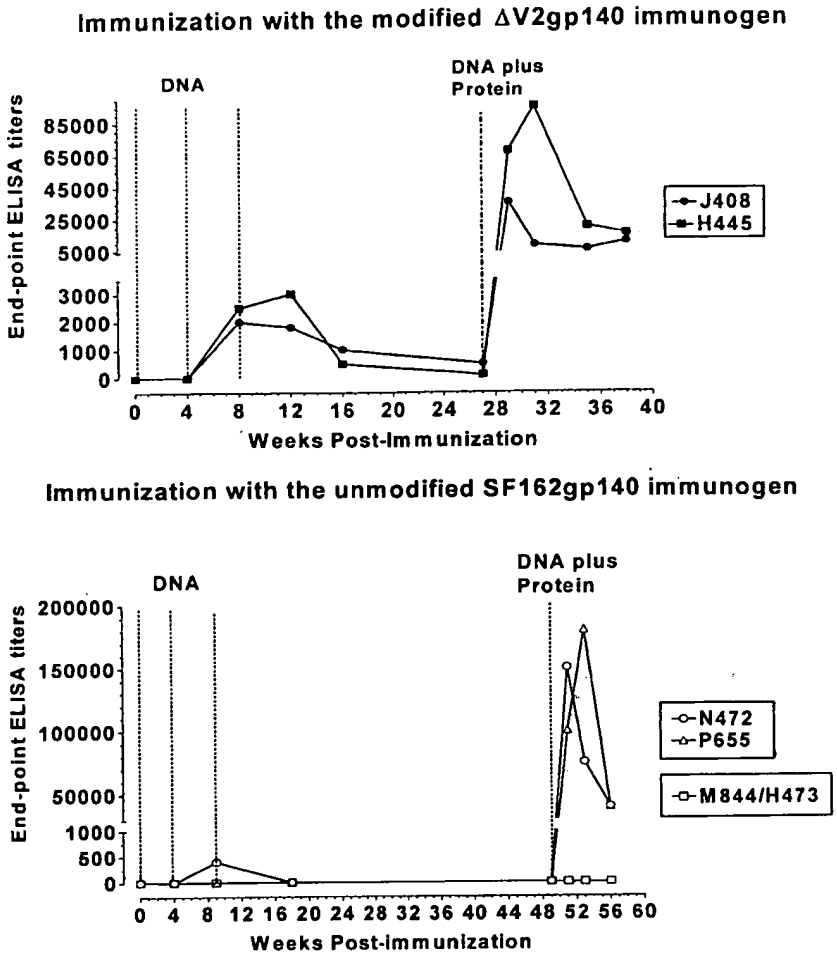


FIG. 3. Generation of antibodies in rhesus macaques. The generation of antienvelope antibodies in two animals (J408 and H445) immunized with the modified ΔV2gp140 immunogen and two animals (P655 and N472) immunized with the unmodified SF162gp140 immunogen, as well as control animals (M844 and H473) immunized with the DNA vector alone, were determined by ELISA using the corresponding protein as described in Materials and Methods. Dashed lines indicate times of immunizations. DNA, animals received three monthly immunizations with DNA vectors expressing the gp140 form of each immunogen. Control animals received the DNA vector alone. DNA plus Protein, animals received a fourth DNA immunization and at the same time were immunized with the corresponding CHO cell-produced oligomeric gp140 proteins, mixed in the adjuvant MF-59C. Control animals received adjuvant alone.

Generation of anti-V3 loop antibodies in rhesus macaques vaccinated with the modified ΔV2 gp140 immunogen. One explanation for the increase in neutralizing activity against the parental SF162 and homologous SF162ΔV2 viruses and the decrease in neutralizing activity against the heterologous isolates following the second booster immunization is that multiple immunizations with the modified ΔV2 gp140 protein increased the titer of antibodies directed against epitopes that are uniquely (or predominantly) expressed on the SF162 and SF162ΔV2 envelopes. It is conceivable that multiple immunizations with the ΔV2 gp140 protein result in the generation of high titers of anti-V3 loop antibodies. To determine the titers of such antibodies, we used V3 loop peptide-based ELISAs using the SF162/SF162ΔV2-derived V3 loop (Fig. 7). This peptide was recognized by antibodies binding to both linear (447D) (7, 12) and conformational (391-95D) (29) epitopes (Fig. 7A). Although anti-V3 loop antibodies were generated upon immunization of macaques with the modified ΔV2 gp140 immunogen, their titers were much lower than those against the entire envelope (Fig. 7B). In addition, the second booster immunization did not increase the titer of anti-V3 loop antibodies. It should be noted, however, that certain anti-V3 loop antibodies present in the sera of these animals may not interact efficiently with the V3 loop peptide in an ELISA format, while they may bind to their epitopes on the native envelope (23). Additionally, the V3 loop peptide used here does not span the carboxy and amino termini of the V3 loop, and our assay does not detect antibodies targeting these two regions. Thus, a more detailed examination of the epitope specificity of the antibodies elicited by the modified ΔV2 gp140 immunogen is required.

DISCUSSION

In this study we compared the immunogenicities of soluble oligomeric gp140 envelope proteins derived from related neutralization-resistant (SF162) and neutralization-susceptible

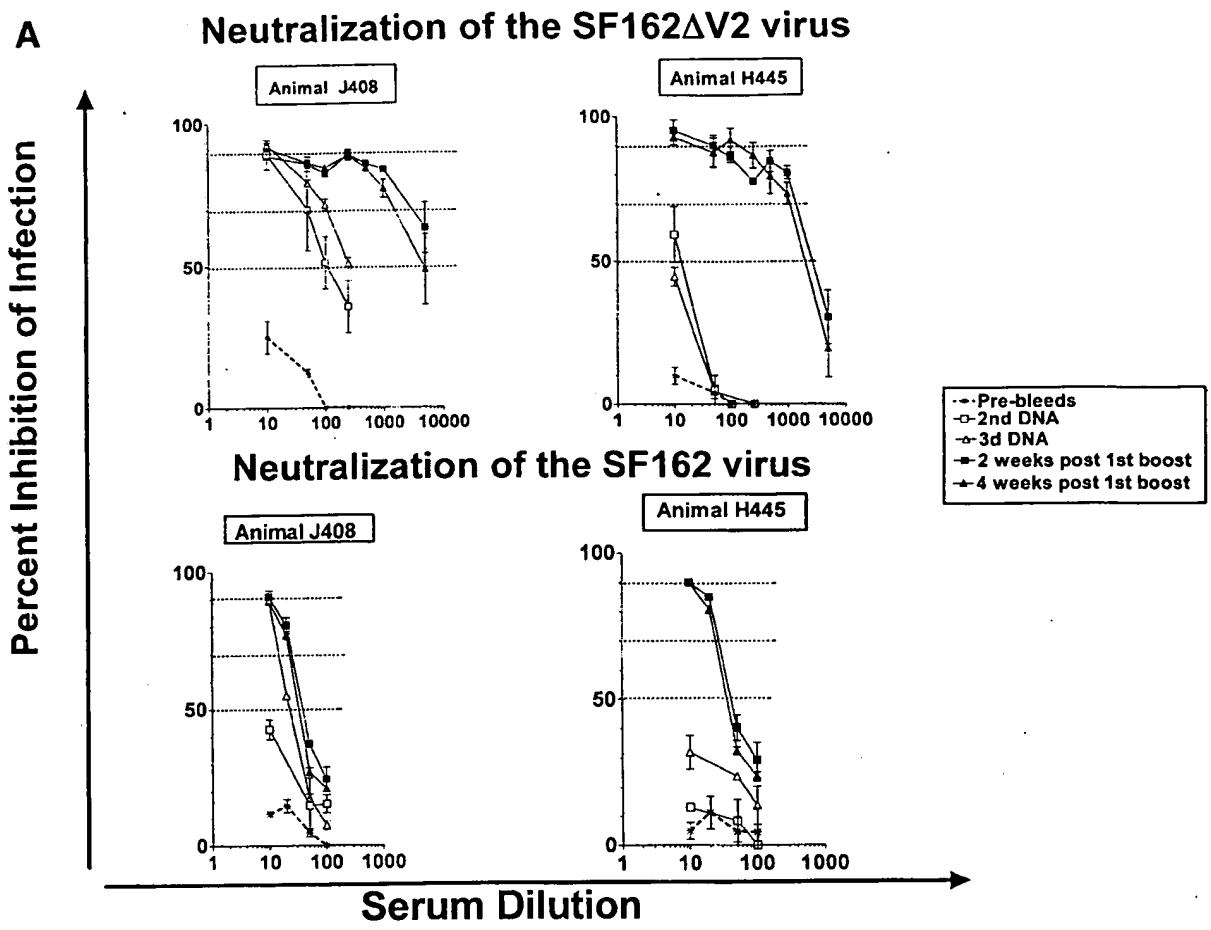


FIG. 4. Neutralizing activities of rhesus macaque sera. Neutralization activities against the SF162 and SF162ΔV2 viruses of sera collected from animals immunized with the modified ΔV2 gp140 (A) and the unmodified (B) SF162 gp140 immunogens were determined as described in Materials and Methods. Dashed lines indicate 50, 70, and 90% inhibition of infection. Results are representative of three to five independent experiments. Data indicate the mean and standard deviation from triplicate wells. Pre-bleeds, sera collected before vaccination; 2nd DNA and 3rd DNA, sera collected 1 month following the second and the third, respectively, DNA administrations; 2 and 4 weeks post 1st boost, sera collected 2 and 4 weeks, respectively, following the DNA-plus-protein booster immunization.

(SF162ΔV2) viruses (30). The only difference between the two immunogens is the absence of 30 amino acids from the V2 loop of the SF162ΔV2-derived immunogen (30).

We first performed immunization studies in rabbits, where we observed that although both proteins elicited similar titers of binding antibodies (Fig. 1), the modified immunogen elicited higher titers of neutralizing antibodies against isolates expressing not only the modified SF162ΔV2 envelope but also the unmodified parental SF162 envelope (Fig. 2). Our results are in agreement with those made on the background of the SIVmac239 virus (25) because they suggest that specifically modified envelope immunogens may elicit antibodies capable of neutralizing isolates expressing the parental unmodified envelope. However, our results contrast with those obtained during the immunization of rabbits with a more extensively modified HIV-1 envelope-derived immunogen, which lacked not only the V2 loop but also the V1 and V3 loops (20). The reasons for this discrepancy are not known, but it would be interesting to determine whether the triple-loop-deleted SF162 envelope also fails to elicit neutralizing antibodies.

In rabbits, both the unmodified SF162 and modified ΔV2 gp140 immunogens elicited neutralizing antibodies against several heterologous primary HIV-1 isolates, but the potential of the modified immunogen to do so was greater (Table 1). Thus, not only did more animals vaccinated with the modified immunogen elicit cross-reactive neutralizing antibodies, but also the breadth and potency of the cross-neutralizing responses were higher in sera collected from these animals than from animals immunized with the unmodified immunogen. We believe that this is a consequence of the greater immunogenicity of conserved neutralizing epitopes on the modified immunogen tested here. Consequently, the modified immunogen more effectively elicits antibodies recognizing these epitopes than the unmodified immunogen. If V2 loop deletion increases the exposure of conserved neutralization epitopes that are poorly exposed on the SF162 envelope, then these two immunogens will elicit different titers of the same antibodies. If V2 loop deletion results in exposure of neutralization epitopes that are not exposed on the SF162 envelope, then the two immunogens will elicit different types of antibodies. An alternative explana-

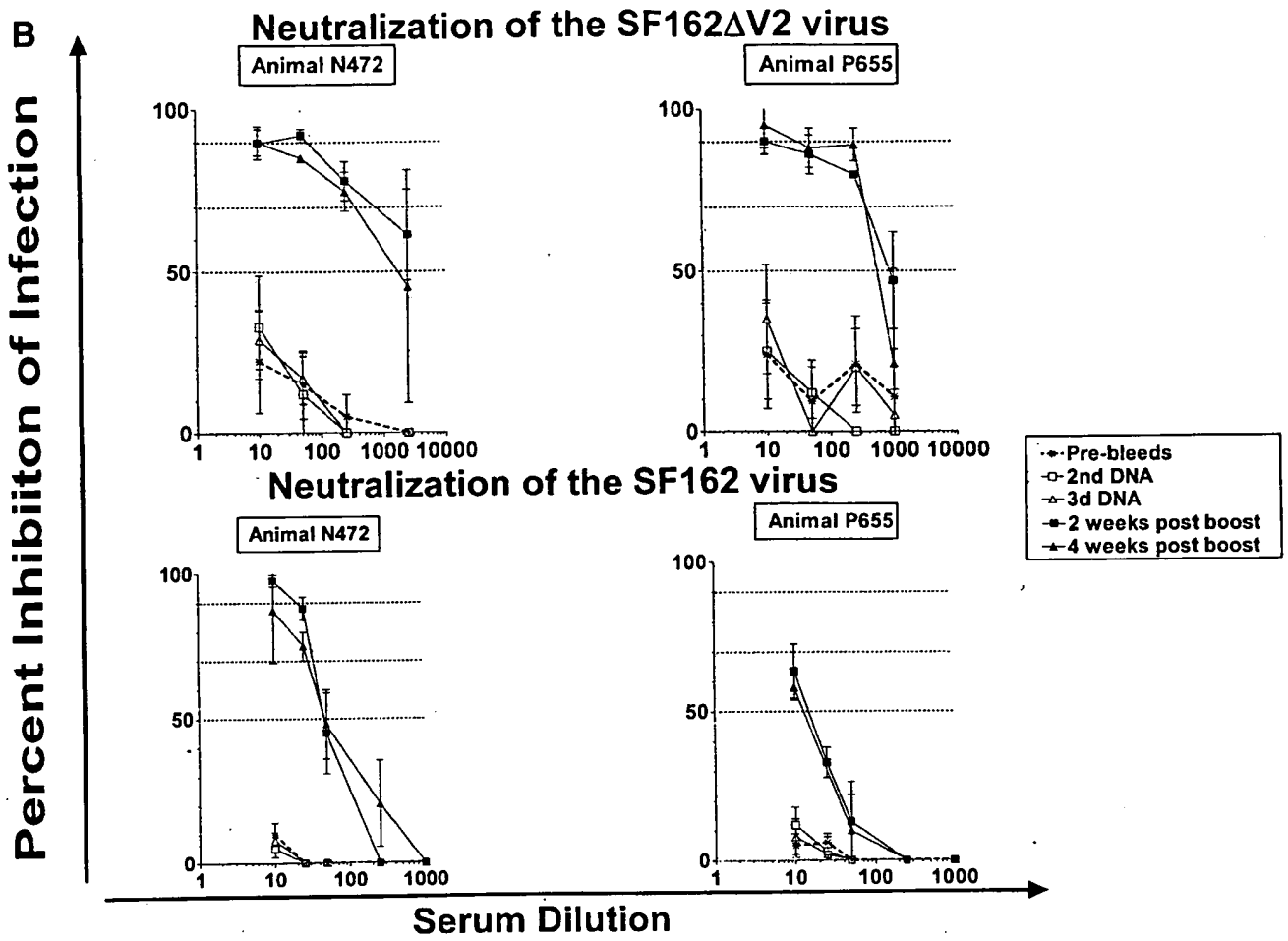


FIG. 4—*Continued.*

tion for the increased ability of the modified ΔV2gp140 immunogen to elicit neutralizing antibodies was previously proposed by our group (30). It is possible that deletion of the V2 loop from the SF162 envelope results in an alteration of the ratio of neutralization and nonneutralization epitopes on this envelope, so that more of the former epitopes are present on the SF162ΔV2 than on the SF162 envelope. If this hypothesis is correct, immunization with the modified ΔV2 gp140 enve-

TABLE 2. Neutralization of heterologous primary HIV-1 isolates by macaque sera

Isolate (coreceptor usage)	MAb		% Neutralization[a] with immunogen[b]							
			ΔV2 gp140				SF162 gp140		SF2 gp120	
			J408		H445					
	2F5	2G12	A	B	A	B	P655 (A)	N472 (A)	L714	L814
91US056(R5)	60	70	90	—	65	—	—	—	—	—
92US714(R5)	70	20	85	—	85	—	—	—	—	—
92US660(R5)	75	70	50	—	80	—	—	—	—	—
92HT593(R5X4)	75	80	—	—	—	—	—	—	—	—
92US657(R5)	NT	NT	—	—	—	65	—	—	—	—
BZ167(R5X4)	90	75	NT	—	NT	80	NT	NT	NT	NT
ADA(R5)	NT	NT	90	50	90	80	NT	NT	NT	NT

[a] Percent neutralization of a given HIV-1 isolate by sera (1:10 dilution) collected from animals immunized with the modified ΔV2 gp140 (J408 and H445), unmodified SF162 gp140 (P655 and N472), and recombinant SF2 gp120 (L714 and L814), calculated as described in Materials and Methods, taking into consideration the nonspecific neutralization recorded with sera collected from the same animals before immunization.

[b] A, sera collected 2 weeks following the DNA-plus-protein booster immunization; B, sera collected 2 weeks following the final protein booster immunization of animals J408 and H445. Values represent averages from two to three independent experiments. The susceptibility of these isolates to neutralization by 2F5 and 2G12 at 25 μg of MAb per ml is also presented. —, not detected; NT, not tested.

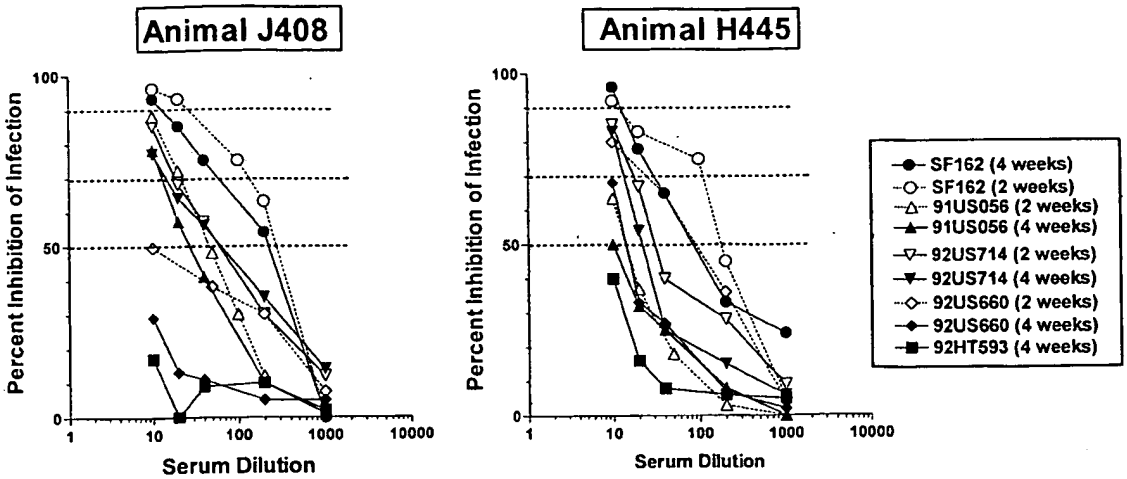


FIG. 5. Neutralization of heterologous clade B primary HIV-1 isolates by macaque sera. Neutralization activities of sera collected 2 and 4 weeks following the DNA-plus-protein booster immunization against isolates heterologous to the vaccine primary HIV-1 isolates were determined as described in Materials and Methods. Dashed lines indicate 50, 70, and 90% inhibition of infection. The values represent specific neutralization, which is defined as the difference between the percent virus neutralization recorded with sera collected following vaccination and that recorded with sera collected prior to the initiation of vaccination. Data points indicate the mean percent specific neutralization from two independent experiments.

lope would be expected to elicit a proportionally higher number of neutralizing antibodies than immunization with the unmodified SF162 gp140 envelope. The above possibilities are not mutually exclusive.

Our vaccination studies conducted in rhesus macaques confirm the observations made in rabbits, that the modified ΔV2 gp140 immunogen is more effective than the unmodified SF162 gp140 in eliciting neutralizing antibodies against isolates expressing the parental SF162 envelope and heterologous envelopes. Neutralization of heterologous primary HIV-1 isolates was less efficient than neutralization of the parental SF162 virus, which in turn was less efficiently neutralized than the homologous SF162ΔV2 virus. This suggests that the epitopes recognized by these cross-reactive antibodies are less accessible on the surface of the heterologous isolates than on the surface of the SF162 virus and much less accessible than they are on the surface of the SF162ΔV2 virus. Based on our previous neutralization studies with MAbs, we believe that deletion of the V2 loop from the SF162 envelope increases the exposure of epitopes participating in envelope-CD4 and -coreceptor binding (30). Such epitopes become only transiently exposed during the displacement of the V1 and V2 loops that take place upon HIV envelope-CD4 binding (37–39). It is possible that these epitopes are normally masked within the oligomeric envelope structure by the V2 loop and deletion of this loop renders them permanently exposed. The difference in neutralization susceptibility between the SF162 and heterologous isolates may be the result of different envelope glycosylation patterns and/or different positioning of the V2 loop on the SF162 and heterologous isolates tested here. Additionally, we expect that although cross-reactive antibodies were elicited by this specific envelope modification, a large fraction of the antibodies elicited by the ΔV2 gp140 immunogen are targeting epitopes unique to the ΔV2 and SF162 envelopes. Our current results suggest that these unique epitopes must be located in envelope regions other than the V3 loop. This is also supported by our observations that the heterologous primary HIV-1 isolates tested here were resistant to neutralization by sera collected from macaques immunized with the recombinant monomeric SF2 gp120 protein, which primarily elicits anti-V3 loop antibodies (36). The lack of generation of high anti-V3 loop antibodies by our soluble oligomeric immunogens may not be due to the specific antigenic structure of these immunogens, because a recent report indicated that immunization with the oligomeric envelope derived from the HXB2 isolate also failed to elicit high titers of V3 loop-directed antibodies (11). Several heterologous primary isolates tested here were, however, completely resistant to neutralization by the antibodies elicited by the modified immunogen. Either these isolates may lack the epitopes recognized by the antibodies elicited by this immunogen or, as mentioned above, these epitopes may be more efficiently masked on these particular isolates than on the isolates susceptible to neutralization. By identifying the epitopes recognized by the neutralizing antibodies elicited by the ΔV2 gp140 immunogen, we may be able to determine whether these epitopes are absent from the heterologous isolates that are resistant to neutralization or whether they are more efficiently occluded on the surface of these isolates.

The envelopes of the SF162 and SF162ΔV2 viruses use the CCR5 cellular protein to mediate virus-cell fusion in the presence of CD4 (30). The heterologous primary isolates tested here used either exclusively CCR5 or both CCR5 and CXCR4 (Table 2). We did not examine whether the antibodies generated by the immunogens tested here, especially the modified one, would have neutralization potential against isolates that exclusively utilize CXCR4 to infect $CD4^+$ T cells. Failure of our immunogens to do so would be an indication that the antibodies elicited may recognize epitopes participating in gp120-CCR5 interaction. Although, we believe that an effec-

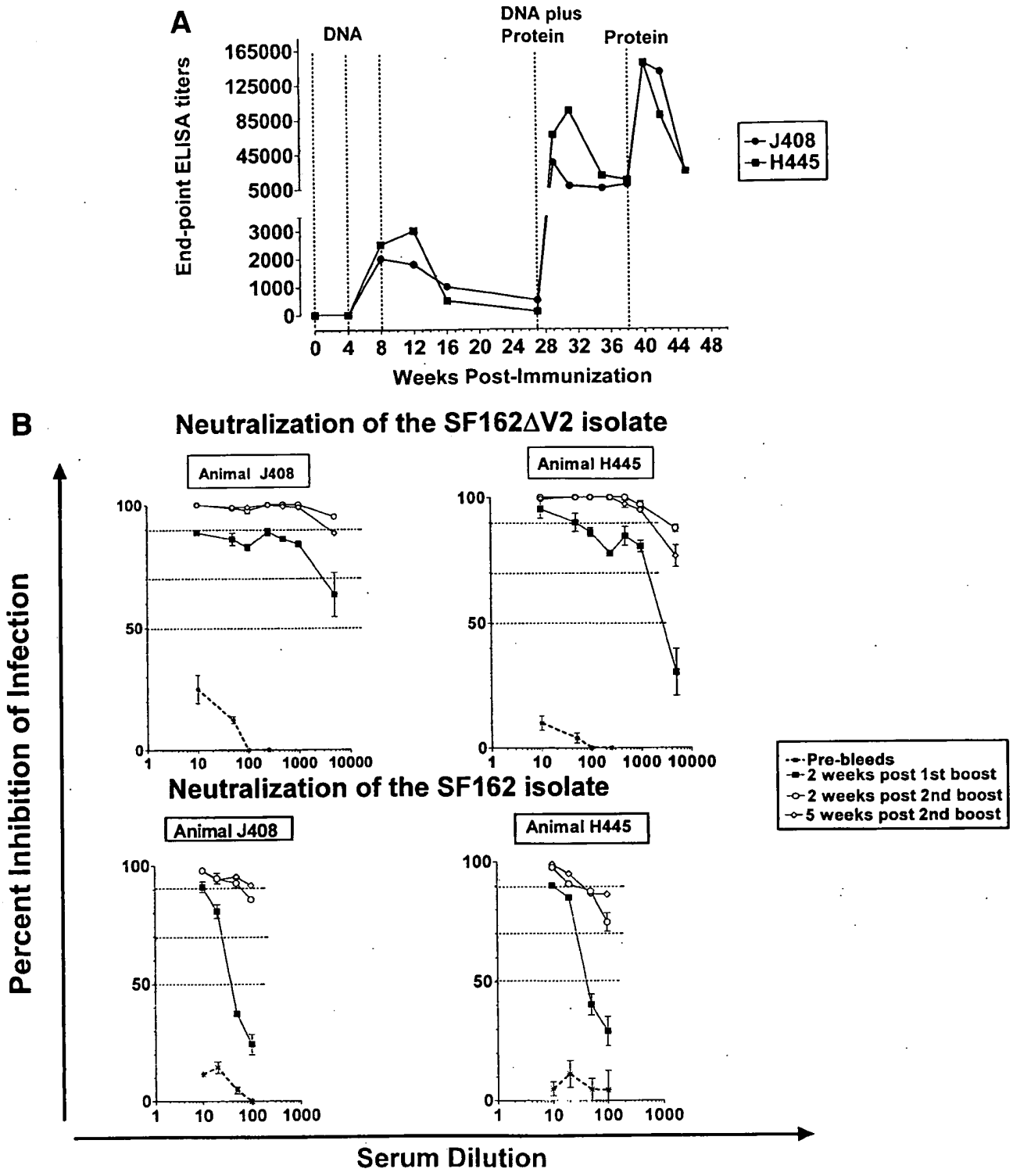


FIG. 6. Generation of binding and neutralizing antibodies following the second booster immunization with the modified ΔV2 gp140 protein. (A) The generation of antienvelope antibodies in two rhesus macaques (J408 and H445) vaccinated with the modified ΔV2 gp140 immunogen was determined by ELISA as described in Materials and Methods. Dashed lines indicate times of immunizations. DNA, animals received three monthly immunizations with DNA vectors expressing the gp140 form of this immunogen; DNA plus Protein, animals received a fourth DNA immunization and purified oligomeric ΔV2 gp140 protein; Protein, the animals were immunized with the purified oligomeric ΔV2 gp140 protein alone. (B) Neutralization activities against the SF162ΔV2 and SF162 isolates of sera following the second boost were compared to those of sera collected following the first boost (see also Fig. 4). Nonspecific neutralization recorded with preimmunization sera (Pre-bleeds) is also shown.

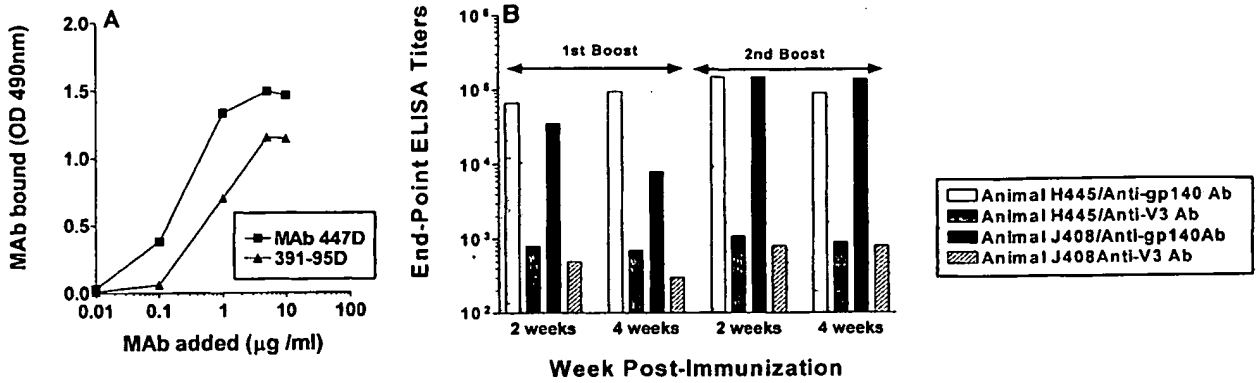


FIG. 7. Presence of anti-V3 loop antibodies in sera collected from macaques immunized with the modified ΔV2 gp140 immunogen. The development of anti-V3 loop antibodies was determined by ELISA using the V3 loop peptide derived from the SF162/SF162ΔV2 envelope. (A) We first examined whether the captured V3 loop peptide interacts with specific anti-V3 loop MAbs recognizing linear (447D) and conformational (391-95D) V3 loop epitopes. (B) We next determined the titer of anti-V3 loop antibodies present in sera collected 2 and 4 weeks following the 1st and 2nd boosts from the two vaccinated animals. As a comparison, we also include the titers of total antienvelope antibodies present in the same sera.

tive vaccine against HIV should elicit a broad spectrum of neutralizing antibodies, such a vaccine should primarily elicit neutralizing antibodies against R5-using HIV-1 isolates because such isolates are more effective than X4-using HIV-1 isolates in establishing a primary HIV-1 infection in exposed humans (15, 19, 26, 42).

Some differences were recorded during the immunization of rabbits and macaques. For example, in rabbits both immunogens elicited similar binding antibody titers, while the unmodified immunogen elicited higher titers of binding antibodies in macaques. Also, although in rabbits the unmodified immunogen elicited neutralizing antibodies (albeit infrequently and at low titers) against a few heterologous HIV-1 isolates, it failed to do so in macaques. Finally, some isolates, such as 92HT593 was neutralized by rabbit but not by macaque sera. Whether these differences are due to differences in the way the immune systems of these two animal species reacts to the same immunogen, or whether they are due to the different methods of immunization (gene gun in the case of rabbits and intramuscular plus intradermal needle DNA injections followed by protein immunization in the case of macaques) merits further investigation.

It is important to note that in the case of the heterologous primary HIV-1 isolates tested here, we rarely recorded 90% inhibition of the infection with the neutralization assay that we used at the lowest serum dilution evaluated (1:10) and that the cross-neutralizing activity was lost within 5 weeks following the DNA-plus-protein booster immunization. We corrected our neutralization data for nonspecific neutralization recorded with autologous sera collected before vaccination, because it was recently reported that preimmunization sera (1:4 dilution) collected from immunized human volunteers could inhibit by up to 80% the infection of several primary HIV-1 isolates (2). Rabbit or macaque preimmunization sera (1:10 dilution) used here generally did not inhibit by over 30 to 35% the infection of the isolates tested. Although it is possible that 90% inhibition of infection may be recorded more frequently at lower serum dilutions, we believe that to increase the potency and breadth of neutralization, additional envelope modifications must be introduced. These modifications should increase the exposure and/or the number of conserved neutralization epitopes on the immunogen. However, because an increase in epitope exposure may not automatically result in an increase of epitope immunogenicity (20), the effect of a particular envelope modification on envelope immunogenicity can be determined empirically only by in vivo immunogenicity studies in relevant animal models. Nevertheless, the achievement of the observed level of cross-reactive neutralizing activity in a nonhuman primate warrants further evaluation of this particular envelope modification approach. Neutralization epitopes may also become exposed during the propagation of HIV in cells lacking receptor molecules (14, 17) or by thermal or chemical treatment of virions (13). Envelope molecules derived from such viruses, when used as immunogens, may also elicit high titers of neutralizing antibodies.

It would also be important to compare the protective potentials of antibodies elicited by our modified and unmodified immunogens. In this regard, we recently reported that the antibodies elicited by the modified ΔV2gp140 immunogen offered partial protection from challenged with the related pathogenic SIV-HIV himeric virus $SHIV_{SF162P4}$ (6). However, the ability of the antibodies elicited by our modified immunogen to protect macaques from heterologous viral challenge is not yet known. Importantly, vaccination methodologies that generate and sustain significant titers of cross-reactive neutralizing antibodies must be developed.

During the second booster immunization (protein alone) of macaques with the modified ΔV2 gp140 immunogen, we observed that although an increase in the potency of neutralizing antibody responses was recorded in the case of the homologous and parental viruses (Fig. 6), a parallel increase against heterologous primary HIV-1 isolates was not recorded (Table 2). Thus, multiple immunizations with the purified modified ΔV2 gp140 protein may preferentially increase the titer of antibodies recognizing epitopes that are unique to the SF162 and ΔV2 envelopes. Whether this is related to the fact that the

DNA-expressed protein has an intact gp120-gp41 cleavage site, while this site is absent from the CHO cell-produced protein is not known. It is possible that conserved neutralization epitopes are more efficiently exposed on soluble oligomeric gp140 proteins with an intact gp120-gp41 cleavage site. In that respect, differences in the structures of gp140 proteins derived from the SF162 and ΔV2 envelopes with intact or absent gp120-gp41 cleavage sites were reported by our group (32).

In summary, our current data strongly suggest that unless specific modifications are introduced in soluble oligomeric envelope constructs derived from R5-using, primary HIV-1 isolates, the breadth and potency of the neutralizing responses elicited by such vaccines will be weak. The challenge is to identify additional envelope modifications that would further increase the immunogenicity of neutralization epitopes whose structures are conserved among primary HIV-1 isolates.

ACKNOWLEDGMENTS

This study was supported by NIH grants AI47708-01 (L.S.), AI44309-01 (L.S.), AI44309 (S.L.), and AI40337 (S.L.).

We acknowledge Cecilia Cheng-Mayer and James Bradac for many helpful discussions throughout this study and John Donnelly for critical reading of the manuscript.

REFERENCES

1. **Bou-Habib, D. C., G. Roderiquez, T. Oravesz, P. W. Berman, P. Lusso, and M. A. Norcross.** 1994. Cryptic nature of envelope V3 region epitopes protects primary monocytotropic human immunodeficiency virus type 1 from antibody neutralization. J. Virol. 68:6006–6013.
2. **Bures, R., A. Gaitan, T. Zhu, C. Graziosi, K. M. McGrath, J. Tartaglia, P. Caudrelier, R. El Habib, M. Klein, A. Lazzarin, D. M. Stablein, M. Deers, L. Corey, M. L. Greenberg, D. H. Schwartz, and D. C. Montefiori.** 2000. Immunization with recombinant canarypox vectors expressing membrane-anchored glycoprotein 120 followed by glycoprotein 160 boosting fails to generate antibodies that neutralize R5 primary isolates of human immunodeficiency virus type 1. AIDS Res. Hum. Retroviruses 16:2019–2035.
3. **Cao, J., N. Sullivan, E. Desjardin, C. Parolin, J. Robinson, R. Wyatt, and J. Sodroski.** 1997. Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. J. Virol. 71:9808–9812.
4. **Chapman, B. S., R. M. Thayer, K. A. Vincent, and N. L. Haigwood.** 1991. Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. Nucleic Acids Res. 19:3979–3986.
5. **Cheng-Mayer, C., M. Quiroga, J. W. Tung, D. Dina, and J. A. Levy.** 1990. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. J. Virol. 64:4390–4398.
6. **Cherpelis, S., I. Shrivastava, A. Gettie, X. Jin, H. D. D, S. W. Barnett, and L. Stamatatos.** 2001. DNA vaccination with the human immunodeficiency virus type 1 SF162ΔV2 envelope elicits immune responses that offer partial protection from simian/human immunodeficiency virus infection to $CD8^+$ T-cell-depleted rhesus macaques. J. Virol. 75:1547–1550.
7. **Conley, A. J., M. K. Gorny, J. A. Kessler, II, L. J. Boots, M. Ossorio-Castro, S. Koenig, D. W. Lineberger, E. A. Emini, C. Williams, and S. Zolla-Pazner.** 1994. Neutralization of primary human immunodeficiency virus type 1 isolates by the broadly reactive anti-V3 monoclonal antibody, 447-52D. J. Virol. 68:6994–7000.
8. **D'Souza, M. P., D. Livnat, J. A. Bradac, and S. H. Bridges.** 1997. Evaluation of monoclonal antibodies to human immunodeficiency virus type 1 primary isolates by neutralization assays: performance criteria for selecting candidate antibodies for clinical trials. AIDS Clinical Trials Group Antibody Selection Working Group. J. Infect. Dis. 175:1056–1062.
9. **Earl, P. L., S. Koenig, and B. Moss.** 1991. Biological and immunological properties of human immunodeficiency virus type 1 envelope glycoprotein: analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. J Virol, 65:31–41.
10. **Earl, P. L., and B. Moss.** 1993. Mutational analysis of the assembly domain of the HIV-1 envelope glycoprotein. AIDS Res. Hum. Retroviruses 9:589–594.
11. **Earl, P. L., W. Sugiura, D. C. Montefiori, C. C. Broder, S. A. Lee, C. Wild, J. Lifson, and B. Moss.** 2001. Immunogenicity and protective efficacy of oligomeric human immunodeficiency virus type 1 gp140. J. Virol. 75:645–653.
12. **Gorny, M. K., A. J. Conley, S. Karwowska, A. Buchbinder, J.-Y. Xu, E. A. Emini, S. Koenig, and S. Zolla-Pazner.** 1992. Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody. J. Virol. 66:7538–7542.
13. **Grovit-Ferbas, K., J. F. Hsu, J. Ferbas, V. Gudeman, and I. S. Chen.** 2000. Enhanced binding of antibodies to neutralization epitopes following thermal and chemical inactivation of human immunodeficiency virus type 1. J. Virol. 74:5802–5809.
14. **Hoffman, T. L., C. C. LaBranche, W. Zhang, G. Canziani, J. Robinson, I. Chaiken, J. A. Hoxie, and R. W. Doms.** 1999. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. Proc. Natl. Acad. Sci. USA 96:6359–6364.
15. **Huang, Y., W. A. Paxton, S. M. Wolinsky, A. U. Neumann, L. Zhang, T. He, S. Kang, D. Ceradini, Z. Jin, K. Yazdanbakhsh, K. Kunstman, D. Erickson, E. Dragon, N. R. Landau, J. Phair, D. D. Ho, and R. A. Koup.** 1996. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat. Med. 2:1240–1243.
16. **Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659.
17. **LaBranche, C. C., T. L. Hoffman, J. Romano, B. S. Haggarty, T. G. Edwards, T. J. Matthews, R. W. Doms, and J. A. Hoxie.** 1999. Determinants of CD4 independence for a human immunodeficiency virus type 1 variant map outside regions required for coreceptor specificity. J. Virol. 73:10310–10319.
18. **LaCasse, R. A., K. E. Follis, M. Trahey, J. D. Scarborough, D. R. Littman, and J. H. Nunberg.** 1999. Fusion-competent vaccines: broad neutralization of primary isolates of HIV. Science 283:357–362.
19. **Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau.** 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86:367–377.
20. **Lu, S., R. Wyatt, J. F. L. Richmond, F. Mustafa, S. Wang, J. Weng, D. C. Montefiori, J. Sodroski, and H. L. Robinson.** 1998. Immunogenicity of DNA vaccines expressing human immunodeficiency virus type 1 envelope glycoprotein with and without deletions in the V1/V2 and V3 regions. AIDS Res. Hum. Retroviruses 14:151–155.
21. **Mascola, J. R., M. G. Lewis, G. Stiegler, D. Harris, T. C. VanCott, D. Hayes, M. K. Louder, C. R. Brown, C. V. Sapan, S. S. Frankel, Y. Lu, M. L. Robb, H. Katinger, and D. L. Birx.** 1999. Protection of macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. J. Virol. 73:4009–4018.
22. **Mascola, J. R., S. W. Snyder, O. S. Weislow, S. M. Belay, R. B. Belshe, D. H. Schwartz, M. L. Clements, R. Dolin, B. S. Graham, G. J. Gorse, M. C. Keefer, M. J. McElrath, M. C. Walker, K. F. Wagner, J. G. McNeil, F. E. McCutchan, D. S. Burke, and the NIAID AIDS Vaccine Evaluation Group.** 1996. Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. J. Infect. Dis. 173:340–348.
23. **Moore, J. P.** 1993. The reactivities of HIV-1+ human sera with solid-phase V3 loop peptides can be poor predictors of their reactivities with V3 loops on native gp120 molecules. AIDS Res. Hum. Retroviruses 9:209–219.
24. **Moore, J. P., J. A. McKeating, Y. Huang, A. Askenazi, and D. D. Ho.** 1992. Virions of primary human immunodeficiency virus type 1 isolates resistant to soluble CD4 (sCD4) neutralization differ in sCD4 binding and glycoprotein gp120 retention from sCD4-sensitive isolates. J. Virol. 66:235–243.
25. **Reitter, J. N., R. E. Means, and R. C. Desrosiers.** 1998. A role for carbohydrates in immune evasion in AIDS. Nat. Med. 4:679–684.
26. **Samson, M., F. Libert, B. J. Doranz, J. Rucker, C. Liesnard, C. M. Farber, S. Saragosti, C. Lapoumeroulie, J. Cognaux, C. Forceille, G. Muyldermans, C. Verhofstede, G. Burtonboy, M. Georges, T. Imai, S. Rana, Y. Yi, R. J. Smyth, R. G. Collman, R. W. Doms, G. Vassart, and M. Parmentier.** 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722–725.
27. **Sattentau, Q. J., and J. P. Moore.** 1991. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. J. Exp. Med. 174:407–415.
28. **Sattentau, Q. J., J. P. Moore, F. Vignaux, F. Traincard, and P. Poignard.** 1993. Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. J. Virol. 67:7383–7393.
29. **Seligman, S. J., J. M. Binley, M. K. Gorny, D. R. Burton, S. Zolla-Pazner, and K. A. Sokolowski.** 1996. Characterization by serial competition ELISAs of HIV-1 V3 loop epitopes recognized by monoclonal antibodies. Mol. Immunol. 33:737–745.
30. **Stamatatos, L., and C. Cheng-Mayer.** 1998. An envelope modification that renders a primary, neutralization resistant, clade B HIV-1 isolate highly susceptible to neutralization by sera from other clades. J. Virol. 72:7840–7845.
31. **Stamatatos, L., and C. Cheng-Mayer.** 1995. Structural modulations of the envelope gp120 glycoprotein of human immunodeficiency virus type 1 upon oligomerization and differential V3 loop epitope exposure of isolates dis-

playing distinct tropism upon virion-soluble receptor binding. J. Virol. 69: 6191–6198.
32. Stamatatos, L., M. Lim, and C. Cheng-Mayer. 2000. Generation and structural analysis of soluble oligomeric envelope proteins derived from neutralization-resistant and neutralization-susceptible primary HIV-1 isolates. AIDS Res. Hum. Retroviruses 16:981–994.
33. Stamatatos, L., M. Wiskerchen, and C. Cheng-Mayer. 1998. Effect of major deletions in the V1 and V2 loops of a macrophage-tropic HIV-1 isolate on viral envelope structure, cell-entry and replication. AIDS Res. Hum. Retroviruses 14:1129–1139.
34. Stamatatos, L., S. Zolla-Pazner, M. Gorny, and C. Cheng-Mayer. 1997. Binding of antibodies to virion-associated gp120 molecules of primary-like human immunodeficiency virus type 1 (HIV-1) isolates: effect on HIV-1 infection of macrophages and peripheral blood mononuclear cells. Virology 229:360–369.
35. Sullivan, N., Y. Sun, J. Li, W. Hofmann, and J. Sodroski. 1995. Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T-cell line-passaged human immunodeficiency virus type 1 isolates. J. Virol. 69:4413–4422.
36. VanCott, T. C., J. R. Mascola, L. D. Loomis-Price, F. Sinangil, N. Zitomersky, J. McNeil, M. L. Robb, D. L. Birx, and S. Barnett. 1999. Cross-subtype neutralizing antibodies induced in baboons by a subtype E gp120 immunogen based on an R5 primary human immunodeficiency virus type 1 envelope. J. Virol. 73:4640–4650.
37. Wyatt, R., P. D. Kwong, E. Desjardins, R. W. Sweet, J. Robinson, W. A. Hendrickson, and J. G. Sodroski. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393:705–711.
38. Wyatt, R., J. Moore, M. Accola, E. Desjardin, J. Robinson, and J. Sodroski. 1995. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. J. Virol. 69:5723–5733.
39. Wyatt, R., and J. Sodroski. 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280:1884–1888.
40. Wyatt, R., N. Sullivan, M. Thali, H. Repke, D. Ho, J. Robinson, M. Posner, and J. Sodroski. 1993. Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. J. Virol. 67:4557–4565.
41. Yang, X., R. Wyatt, and J. Sodroski. 2001. Improved elicitation of neutralizing antibodies against primary human immunodeficiency viruses by soluble stabilized envelope glycoprotein trimers. J. Virol. 75:1165–1171.
42. Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Coa, R. A. Koup, and D. D. Ho. 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science 261:1179–1181.
43. zur Megede, J., M. C. Chen, B. Doe, M. Schaefer, C. E. Greer, M. Selby, G. R. Otten, and S. W. Barnett. 2000. Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 *gag* gene. J. Virol. 74:2628–2635.

JOURNAL OF VIROLOGY, Feb. 2001, p. 1547–1550
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.3.1547–1550.2001

Vol. 75, No. 3

DNA Vaccination with the Human Immunodeficiency Virus Type 1 SF162ΔV2 Envelope Elicits Immune Responses That Offer Partial Protection from Simian/Human Immunodeficiency Virus Infection to CD8$^+$ T-Cell-Depleted Rhesus Macaques

S. CHERPELIS,[1] I. SHRIVASTAVA,[2] A. GETTIE,[1,3] X. JIN,[1] D. D. HO,[1] S. W. BARNETT,[2] AND L. STAMATATOS[1]*

Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10021-6399[1]; Chiron Corporation, Emeryville, California 94608-2916[2]; and Tulane Regional Primate Research Center, Covington, Louisiana 70433

Received 14 July 2000/Accepted 30 October 2000

DNA immunization of macaques with the SF162ΔV2 envelope elicited lymphoproliferative responses and potent neutralizing antibodies. The animals were depleted of their CD8$^+$ T lymphocytes and then challenged intravenously with SHIV162P4. Compared to unvaccinated animals, the vaccinated macaques had lower peak viremia levels, rapidly cleared plasma virus, and showed delayed seroconversion.

DNA immunization stimulates both the cellular and humoral arms of the immune system (9, 16–18) and elicits immune responses capable of preventing infection of animals by slowly replicating viruses, such as human immunodeficiency virus type 1 (HIV-1) in chimpanzees (3). However, when the challenge virus replicates efficiently in the host, such as simian immunodeficiency virus (SIV) or SIV/HIV (SHIV) in macaques, the DNA-elicited immune responses offer only partial protection (4, 10, 14). To increase the potency of these responses, especially the development of high anti-HIV or -SIV envelope antibody titers, follow-up administration of soluble viral envelope proteins, viral particles or recombinant vaccinia-based viruses expressing the HIV or SIV envelope is required (1, 2, 8, 12–14). This bimodal method of immunization elicits responses capable of protecting rhesus macaques from infection by highly replication-competent SHIV (8, 14). It is unclear, however, whether the recorded protection was mediated by the cellular and/or humoral antiviral responses elicited during DNA immunization. By evaluating and comparing the respective antiviral protective roles of these two types of responses, we hope to develop more effective DNA immunization protocols.

Two rhesus macaques (H445 and J408) were immunized both intradermally and intramuscularly at weeks 0, 4, and 8 with a DNA vector (5, 22) (2 mg of total DNA each time) expressing the SF162ΔV2 gp140 envelope with an intact gp120-gp41 cleavage site (21). The DNA construct was codon optimized for high expression in mammalian cells. At week 27, the animals were immunized one additional time with DNA and with the CHO-produced, purified oligomeric SF162ΔV2 gp140 protein (100 μg) mixed with the MF-59C adjuvant. At week 38, the animals were immunized one additional time with the adjuvanted protein alone.

* Corresponding author. Mailing address: Aaron Diamond AIDS Research Center, 455 First Ave., 7th floor, New York, NY 10016. Phone: (212) 448-5000. Fax: (212) 725-1126. E-mail: 1stamata@adarc.org.

The development of binding antibodies was evaluated by enzyme-linked immunosorbent assay (ELISA) methodologies (20). Antibodies were detectable following the second DNA immunization, and their titers did not increase following the third DNA immunization (Fig. 1A). During the following 5 months, the titers decreased gradually, but were always detectable. The first "boost" increased the titers by approximately 1 to 2 $\log_{10}$ from the peak value recorded following the third DNA immunization. The titers gradually decreased and leveled off during the following 11 weeks, at which point the animals received a second boost, which further increased the antibody titers. Neutralizing antibodies (NA) were evaluated by using the activated peripheral blood mononuclear cell (PBMC) target assay (19), using preimmunization sera to correct for nonspecific neutralization (Fig. 1B). Following the third DNA immunization, the NA titers in animal H445 were lower than those in animal J408, even though the binding antibody titers were similar between the two animals. The NA titers against both SF162ΔV2 and SF162 increased significantly during the subsequent boosts. Vaccine-specific proliferative responses were recorded in both animals. Stimulation indices (SIs) of 5 and 10 were recorded following the first boost in animals J408 and H445, respectively. The second boost increased the potency of these responses in animal H445 (SI of 25), but not in animal J408 (SI of 5).

To evaluate the protective role of the anti-HIV envelope antibodies elicited by our vaccine, we depleted the CD8$^+$ T lymphocytes from the vaccinated animals prior to viral challenge (Fig. 2). CD8 depletion was achieved by three intravenous administrations of the anti-CD8 MAb OKT8F (2 mg/kg of body weight) at daily intervals (7). CD8$^+$ T lymphocytes remained undetectable for approximately 10 days. Concomitantly, we recorded a decrease in the total number of circulating CD3$^+$ T cells. This indicates that the recorded depletion of CD8$^+$ T cells from the periphery is due to their actual elimination. Although we did not evaluate CD8 depletion from the lymph nodes, it was previously demonstrated that a concomitant depletion of CD8$^+$ T cells from the periphery and lymph

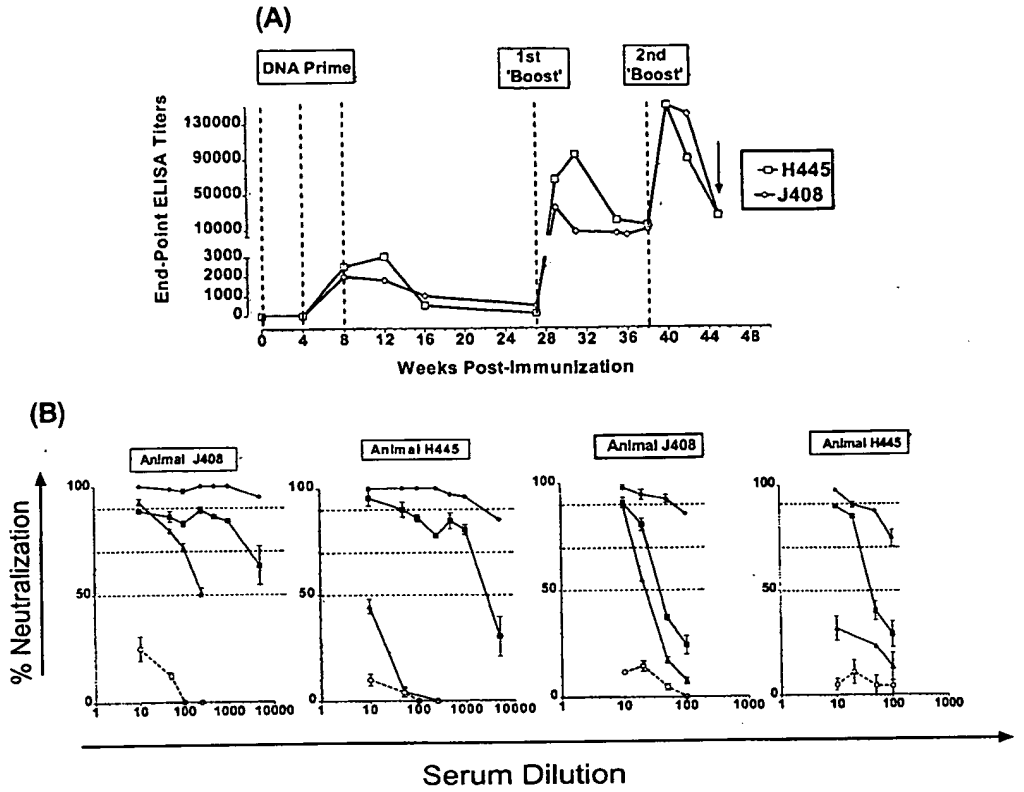


FIG. 1. Generation of anti-HIV envelope antibodies during immunization. (A) Binding antibodies. The envelope-specific titers of binding antibodies in animals J408 and H445 throughout our immunization schedule were determined against the vaccine, i.e., the purified oligomeric SF162ΔV2 gp140 protein. Dashed lines indicate the time of immunization and the arrow indicates the time of viral challenge. (B) Neutralizing antibodies. The presence of neutralizing antibodies against the homologous SF162ΔV2 virus (left panels) and the parental SF162 viruses (right panels) was determined at various time points during the immunization schedule: O, prebleeds; ▲, 1 month past the third DNA immunization; ■, 2 weeks following the 1st boost; and ◆, 2 weeks following the 2nd boost.

nodes occurs when anti-CD8 monoclonal antibodies (MAbs) are introduced into the blood circulation of macaques (11, 15).

One day following the last administration of OKT8F, the immunized and two unimmunized naive animals were challenged intravenously with 100 50% tissue culture infective doses of a cell-free stock of the SHIV162P4 virus (6). This isolate was neutralized by 50 and 90% by sera (1:5 dilution) collected at the day of challenge from animals H445 and J408, respectively.

Both vaccinated and unvaccinated animals became infected; however, differences in the peak viral load levels and viral set points were noted between the two groups (Fig. 3A). Eleven days postchallenge, plasma viremia in the vaccinated animal H445 was lower by 2 and 4 $\log_{10}$ compared to that of the unvaccinated animals A141 and AT54, respectively, while the vaccinated animal J408 was aviremic. At peak viremia, plasma viral levels in the vaccinated animals were 1 to 4 $\log_{10}$ lower than those in the unvaccinated animals. Following peak viremia, an initial rapid decrease followed by a more gradual decrease in plasma viral loads was recorded in the unvaccinated animal A141, while sustained high viral loads were recorded in the second unvaccinated animal, AT54. A very rapid decrease to undetectable levels was recorded in both vaccinated animals within 35 days postchallenge.

Concomitant with the appearance of plasma viremia in the vaccinated animal H445, a rapid increase (by approximately fivefold) of the anti-HIV envelope antibody titers was monitored (Fig. 3B). Subsequently, as the viral load in this animal decreased to undetectable levels, the antibody titers gradually decreased to prechallenge titers. In contrast, the antienvelope antibody titers did not increase in the second vaccinated animal, J408, which had the lowest levels of peak plasma viremia. In the unvaccinated animals, anti-HIV envelope antibodies became detectable approximately 30 days postchallenge. Although their titers increased over time in animal A141 they remained weak and eventually declined prior to death in animal AT54.

The two unvaccinated animals seroconverted to SIV $p27^{gag}$ and pol 31 proteins within 2 weeks postchallenge, while the two vaccinated animals remained seronegative for the first 17 weeks postchallenge (data not shown). Also, although virus was recoverable from rhesus macaque PBMCs collected from the unvaccinated animals at 18, 42, and 48 days postchallenge, it was only recoverable at day 18 from the vaccinated animals (data not shown). Finally, in contrast to the two vaccinated animals and the unvaccinated animal A141, which have remained healthy so far, the second unvaccinated animal, AT54, died from simian AIDS 16 weeks postchallenge.

We believe that the anti-HIV envelope antibodies elicited by our vaccination, although not capable of completely eliminat-

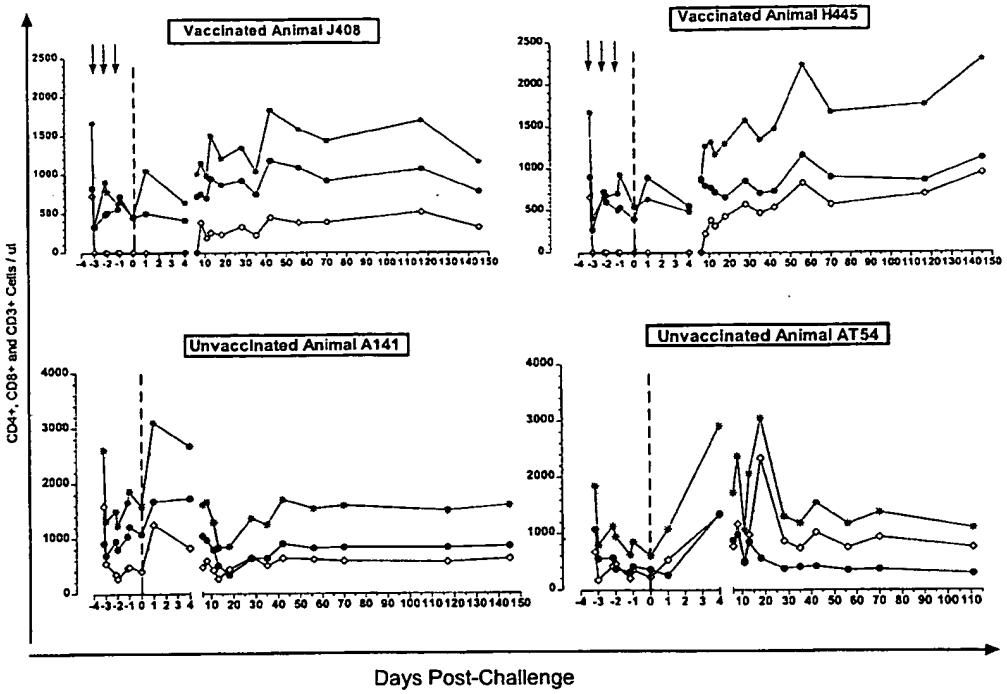


FIG. 2. Depletion of $CD8^+$ T lymphocytes. $CD8^+$ T lymphocytes were depleted from the vaccinated animals by bolus injection of the anti-CD8 MAb OKT8F (arrows). The numbers of circulating $CD4^+$ (solid symbols), $CD8^+$ T (open symbols), and total $CD3^+$ T lymphocytes (asterisks) from vaccinated and unvaccinated animals was determined in samples collected at various points prior to and following SHIV162P4 challenge (dashed line).

ing all the infectious SHIV162P4 viral particles introduced in the blood circulation during challenge, reduced their number, so that fewer numbers of $CD4^+$ T lymphocytes became initially infected in the vaccinated than the unvaccinated animals. This in turn resulted in lower peak levels of plasma viremia during acute infection in the vaccinated animals. The observation that the lowest levels of peak plasma viremia were recorded in the vaccinated animal J408, whose serum had the strongest neutralizing activity against SHIV162P4 at the day of challenge, suggests that neutralizing antibodies played an important protective role during the first 7 days postchallenge. However, in addition to neutralizing antibodies, envelope-specific antibod-

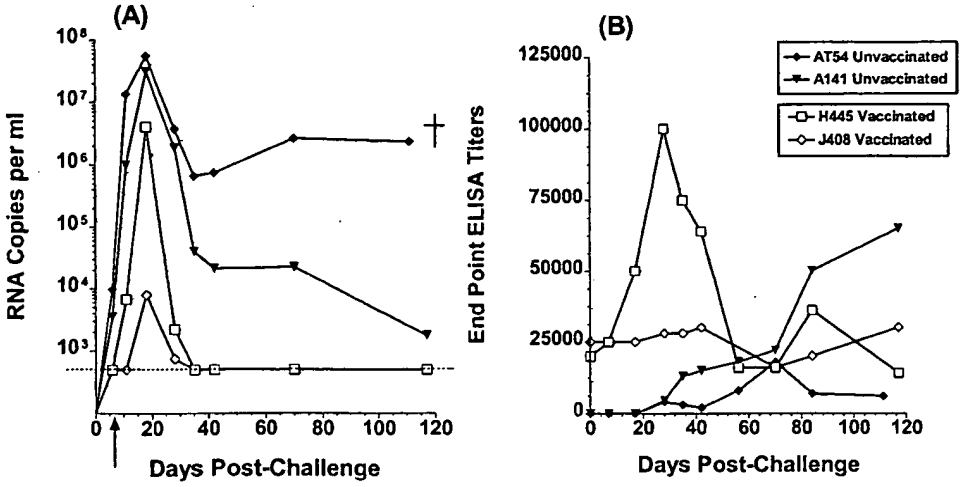


FIG. 3. Viral load and generation anti-HIV envelope antibody titers following SHIV162P4 exposure. (A) Viral load expressed as number of RNA copies per milliliter of plasma. Dashed lines indicate the detection limit of this assay (<500 copies per ml). †, animal AT54 was euthanized 111 days postchallenge following the development of simian AIDS. The arrow indicates the time at which $CD8^+$ T lymphocytes reappeared in the periphery of the vaccinated animals. (B) The generation of anti-HIV envelope antibodies following SHIV162P4 challenge was monitored by SF162ΔV2 gp140-based ELISA methodology. The end point ELISA titers are presented.

ies without neutralizing activity may have been elicited by our vaccine and may also have contributed to viral clearance. The fact that strong anamnestic anti-HIV envelope responses were developed in animal H445 immediately following SHIV challenge indicates that antibodies contributed to the rapid viral clearance to undetectable levels. However, because the $CD8^+$ lymphocytes reappeared in the periphery of the vaccinated animals 7 days postchallenge, they may also have contributed to this rapid viral clearance.

We should point out, however, that in addition to protective antibody responses, our vaccination might have elicited $CD4^+$ T-cell-mediated protective responses, which most likely were present during CD8 depletion. It would be important to assess whether DNA vaccination elicits such responses.

In summary, these pilot studies highlight the important protective role of non-CD8-mediated DNA-vaccine-induced anti-HIV envelope responses and strongly suggest that efforts to develop an effective anti-HIV vaccine should not only aim at eliciting strong cytotoxic T-lymphocyte responses.

This study was supported in part by NIH grant AI 42670-01 (L.S.). We are grateful to T. Mercalino from R. W. Johnson Pharmaceutical Research Institute for providing the anti-CD8 MAb OKT8F and to Jenny Booth at the Bayer Reference Testing Laboratory for performing the b-DNA assays. We thank A. Ly, Y. Liang, S. Hilt, and A. Fong for technical assistance. We acknowledge Cecilia Cheng-Mayer and Janet Harouse (ADARC) for many helpful discussions throughout this study and for providing the SHIV162P4 virus.

REFERENCES

1. **Agadjanyan, M. G., N. N. Trivedi, S. Kudchodkar, M. Bennett, W. Levine, A. Lin, J. Boyer, D. Levy, K. E. Ugen, J. J. Kim, and D. B. Weiner.** 1997. An HIV type 2 DNA vaccine induces cross-reactive immune responses against HIV type 2 and SIV. AIDS Res. Hum. Retrovir. **13**:1561–1572.
2. **Barnett, S. W., J. M. Klinger, B. Doe, C. M. Walker, L. Hansen, A. M. Duliege, and F. M. Sinangil.** 1998. Prime-boost immunization strategies against HIV. AID Res. Hum. Retrovir. **14**(Suppl. 3):S299–S309.
3. **Boyer, J. D., K. E. Ugen, B. Wang, M. Agadjanyan, L. Gilbet, M. L. Bagarazzi, M. Chattergoon, P. Frost, A. Javadian, W. V. Williams, Y. Refaeli, R. B. Ciccarelli, D. McCallus, L. Coney, and D. B. Weiner.** 1997. Protection of chimpanzees from high-dose heterologous HIV-1 challenge by DNA vaccination. Nat. Med. **3**:526–532.
4. **Boyer, J. D., B. Wang, K. E. Ugen, M. Agadjanyan, A. Javadian, P. Frost, K. Dang, R. A. Carrano, R. Ciccarelli, L. Coney, W. V. Williams, and D. B. Weiner.** 1996. In vivo protective anti-HIV immune responses in non-human primates through DNA immunization. J. Med. Primatol. **25**:242–250.
5. **Chapman, B. S., R. M. Thayer, K. A. Vincent, and N. L. Haigwood.** 1991. Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. Nucleic Acids Res. **19**:3979–3986.
6. **Harouse, J. M., A. Gettie, R. C. Tan, J. Blanchard, and C. Cheng-Mayer.** 1999. Distinct pathogenic sequela in rhesus macaques infected with CCR5 or CXCR4 utilizing SHIVs. Science **284**:816–819.
7. **Jin, X., D. E. Bauer, S. E. Tuttleton, S. Lewin, A. Gettie, J. Blanchard, C. E. Irwin, J. T. Safrit, J. Mittler, L. Weinberger, L. G. Kostrikis, L. Zhang, A. S. Perelson, and D. D. Ho.** 1999. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. J. Exp. Med. **189**:991–998.
8. **Letvin, N. L., D. C. Montefiori, Y. Yasutomi, H. C. Perry, M.-E. Davies, C. Lekutis, M. Alroy, D. L. Freed, C. I. Lord, L. K. Handt, M. A. Liu, and J. W. Shiver.** 1997. Potent protective anti-HIV immune responses generated by bimodal HIV envelope DNA plus protein vaccination. Proc. Natl. Acad. Sci. USA **94**:9378–9383.
9. **Liu, M. A., Y. Yasutomi, M.-E. Davis, H. C. Perry, D. C. Freed, N. L. Letvin, and J. W. Shiver.** 1996. Vaccination of mice and nonhuman primates using HIV-gene-gun-containing DNA, vol. 48. Karger, Basel, Switzerland.
10. **Lu, S., J. Arthos, D. C. Montefiori, Y. Yasutomi, K. Manson, F. Mustafa, E. Johnson, J. C. Santoro, J. Wissink, J. I. Mullins, J. R. Haynes, N. L. Letvin, M. Wyand, and H. L. Robinson.** 1996. Simian immunodeficiency virus DNA vaccine trial in macaques. J. Virol. **70**:3978–3991.
11. **Matano, T., R. Shibata, C. Siemon, M. Connors, H. C. Lane, and M. A. Martin.** 1998. Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. J. Virol. **72**:164–169.
12. **Richmond, J. F. L., S. Lu, J. C. Santoro, J. Weng, S.-L. Hu, D. C. Montefiori, and H. L. Robinson.** 1998. Studies of the neutralizing activity and avidity of antihuman immunodeficiency virus type 1 Env antibody elicited by DNA priming and protein boosting. J. Virol. **72**:9092–9100.
13. **Richmond, J. F. L., F. Mustafa, S. Lu, J. C. Santoro, J. Weng, M. O'Connell, E. M. Fenyo, J. L. Hurwitz, D. C. Montefiori, and H. L. Robinson.** 1997. Screening of HIV-1 Env glycoproteins for the ability to raise neutralizing antibody using DNA immunization and recombinant vaccinia virus boosting. Virology **230**:265–274.
14. **Robinson, H. L., D. C. Montefiori, R. P. Johnson, K. H. Manson, M. L. Kalish, J. D. Lifson, T. A. Rizvi, S. Lu, S. L. Hu, G. P. Mazzara, D. L. Panicali, J. G. Herndon, R. Glickman, M. A. Candido, S. L. Lydy, M. S. Wyand, and H. M. McClure.** 1999. Neutralizing antibody-independent containment of immunodeficiency virus challenges by DNA priming and recombinant pox virus booster immunizations. Nat. Med. **5**:526–534.
15. **Schmitz, J. E., M. J. Kuroda, S. Santra, V. G. Sasseville, M. A. Simon, M. A. Lifton, P. Racz, K. Tenner-Racz, M. Dalesandro, B. J. Scallon, J. Ghrayeb, M. A. Forman, D. C. Montefiori, E. P. Rieber, N. L. Letvin, and K. A. Reimann.** 1999. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science **283**:857–860.
16. **Shiver, J. W., M.-E. Davies, H. C. Perry, D. C. Freed, and M. A. Liu.** 1996. Humoral and cellular immunities elicited by HIV-1 DNA vaccination. J. Pharm. Sci. **85**:1317–1324.
17. **Shiver, J. W., H. C. Perry, M.-E. Davies, D. C. Freed, and M. A. Liu.** 1995. Cytotoxic T lymphocyte and helper T cell responses following HIV polynucleotide vaccination. DNA Vaccines **772**:198–208.
18. **Shiver, J. W., J. B. Ulmer, J. J. Donnely, and M. A. Liu.** 1996. Humoral and cellular immunities elicited by DNA vaccines: application to the human immunodeficiency virus and influenza. Adv. Drug Del. Rev. **21**:19–31.
19. **Stamatatos, L., and C. Cheng-Mayer.** 1998. An envelope modification that renders a primary, neutralization-resistant, clade B human immunodeficiency virus type 1 isolate highly susceptible to neutralization by sera from other clades. J. Virol. **72**:7840–7845.
20. **Stamatatos, L., and C. Cheng-Mayer.** 1995. Structural modulations of the envelope gp120 glycoprotein of human immunodeficiency virus type 1 upon oligomerization and differential V3 loop epitope exposure of isolates displaying distinct tropism upon virion-soluble receptor binding. J. Virol. **69**: 6191–6198.
21. **Stamatatos, L., M. Lim, and C. Cheng-Mayer.** 2000. Generation and structural analysis of soluble oligomeric envelope proteins derived from neutralization-resistant and neutralization-susceptible primary HIV-1 isolates. AIDS Res. Hum. Retrovir. **16**:981–994.
22. **zur Megede, J., M.-C. Chen, B. Doe, M. Schaefer, C. E. Greer, M. Selby, G. R. Otten, and S. W. Barnett.** 2000. Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 *gag* gene. J. Virol. **74**:2628–2635.